UNIVERSITÉ CATHOLIQUE DE LOUVAIN

FACULTÉ DES SCIENCES ÉCONOMIQUES, SOCIALES ET POLITIQUES

Nouvelle série-N°259

TESTING THE LINK SPECIFICATION IN BINARY CHOICE MODELS. A SEMIPARAMETRIC APPROACH.

Isabel PROENÇA

CIACO
Louvain-la-Neuve
1995

OBRA ENTREGUE PARA APRECIAÇÃO DA EQUIVALÊNCIA AO GRAU DE DOUTOR EM ECONOMIA CONCEDIDA PELA UNIVERSIDADE TÉCNICA DE LISBOA ATRAVÉS DO INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO AO ABRIGO DO DECRETO-LEI N.º 283/83, DE 21 DE JUNHO. ESTA OBRA NÃO PODERÁ SER REPRODUZIDA SEM EXPRESSO CONSENTIMENTO DO AUTOR, RESSALVADAS AS DISPOSIÇÕES CONSTANTES DO ART.º 76.º, DO CÓDIGO DO DIREITO DO AUTOR (DECRETO-LEI N.º 63/85, DE 14 DE MARÇO).

LISBOA, 12 DE SETEMBRO DE 1996

O PRESIDENTE DO JÚRI,

Je soussigné, Pierre DEHEZ,
Président du Département des
Sciences Economiques de
l'Université Catholique de Louvain,
certifie que Madame Isabel PROENÇA
a défendu avec succès la thèse de
doctorat, dont le présent ouvrage
reprend l'intégralité du texte,
en date du 27 novembre 1995.

Fait à Louvain-la-Neuve, le 30 novembre
1995

Para as minhas irmãs, Paula e Helena

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply grateful to my parents for their care, encouragement and example which have enlightened me and given a meaning for the effort I had to devote to accomplish this work.

I am special grateful to my dear friends Isabel, Carlos and João Luis who have convinced me and gave me the courage to start this adventure.

I am strongly indebted to Professor Bento Murteira who guided my research in Portugal for all his care and support. His dedication to work, honesty and generosity are for me a permanent inspiration.

This thesis would not have been possible without the commitment of my advisor Wolfgang Härdle. I have learned from him not only about econometrics and research but also about life. He has taught me to be ambitious, self confident in improving my capacities and to fight for what I believe. I am thankful for all his support and the opportunities he gave me.

I feel a deep gratitude to my co–author Christian Ritter. His determination and guidance have encourage me to proceed the work. Without him chapter 6 would not have been possible. I am thankful to my other co–author Axel Werwartz. To work together with them was a pleasure and a rich experience.

I have to acknowledge the members of the jury of this thesis, Joel Horowitz, Luc Bawens, Leopold Simar and Alois Kneip for their valuable comments and suggestions that much have improved this dissertation. All the errors are my responsibility. A special thank is due to Luc Bawens whose careful notes on the original text have being a precious help in its revision.

I am thankful to my friend João Santos Silva who read the manuscript and gave me interesting suggestions. For instances, he has pointed me out the work of Ruud.

I am grateful to all professors and visitors of CORE and Institut de Statistique with whom I have discussed my work and received valuable comments. I have to mention among them Alexander Tsybakov, Alois Kneip, Birgit Grund, Enno Mammen, Hidehiko Ichimura, Joel Horowitz, Jörg Polzehl, Luc Bawens, Michel Delecroix, Peter Hall and Rodney Wolff.

While I was doing this work I have been invited by some researcher centers where I had very fruitful discussions and enjoyed my stay. Wolfgang Härdle has invited me several times to visit the Institut für Statistik und Ökonometrie, Humboldt–Universität zu Berlin, Juan Rodriguez has invited me to the Economics department of the Universidade de Bilbao, Wenceslao González-Manteiga has invited me to the Mathematics department of the Universidade de Santiago de Compostela, Michel Delecroix has invited me to ENSAE, Paris and Emmanuel Jolivet has invited me to INRA, Paris. To them and the respective researcher centers all my gratitude.

I am grateful to Joel Horowitz and to Ludwig Fahrmeir who allowed me to use the data sets about respectively mode choice to travel on the way to work and credit–scoring. I thank Marlene Müler who was helpful on providing me these data sets. I am grateful too to Jean-Marie Rolin who allowed me to use the other credit-scoring data set referring to the Flemish region of Belgium. I thank also Cecile Denis in her help on arranging that data set. Still about data, I have to remark the great work of Axel Werwatz on preparing the data set about unemployment after apprenticeship. All my thanks.

I have to acknowledge the way I was welcomed at CORE and Institut de Statistique, their cooperative staff, and the stimulating scientific environment that made possible the research behind this thesis.

I owe special thanks to Berwin Turlach and Sigbert Klinke for their expert help with computers and computation problems. Berwin gave me also a precious help with LateX. I wish to thank too Valentin Patilea and Christian Weiner for the fruitful discussions and fine insights.

I have to thank the financial support I had that made possible the achievement of this work. Namely, the grant I had from JNICT (Junta Nacional de Investigação Científica) and the grant provided by FDS to pursuit the development of the project "Enseignement da la Statistique sur ordinateur personnel". I have to thank also the support of the Insituto Superior de Economia e Gestão (ISEG), Universidade Técnica de Lisboa which has released me from my teaching duties during my stay at Louvain-La-Neuve.

I am grateful to Heraklis Polemarchakis and Jayasry Dutta for their generousity and the care they show to the Ph. D. students.

During my stay at Louvain-La-Neuve I had the fortune to meet a lot of nice people with whom I have shared pleasant moments, namely my colleagues at CORE and Institut de Statistique. I am thankful for they made my life enjoyable.

Finally, to my friends that made this work possible for being at my side in the bad moments and gave me the joy of their companionship at Louvain-La-Neuve I wish to express all my deep gratitude: Isabel, Cinzia, Marlene, Juan, Maria João, Pedro, Miguel and Macarena, Fátima, Abdur, Manuel, Fernando, Irene, Sigbert, Marco and Marcela, Marga, Arantza, Viviane, and specially Berwin. You will remain all in my heart.

Introduction

This thesis will be centered on the field of microeconometrics by dealing with models that describe the behavior of individual decision making units. Microeconomic theory provides a rich framework to analyze and understand the motive of individual decision units. Econometrics has the task to quantify the structure describing individual decision actions for a given population using data that are real measurements characterizing the individuals of this population. There is surely an important interaction between both disciplines. Econometrics needs the knowledge of microeconomics to specify the models quantifying the individual behavior. However, econometrics results are important to economic theory to validate the assumptions and hypotheses postulated or to guide the specification of new hypotheses.

The models discussed in this thesis belong to the class of qualitative choice models (also known as discrete choice models). These models specify the probability that a given individual will choose a particular alternative from a well specified set of alternatives. Probit, logit, conditional probit and multinomial logit are examples of such models. There are many applications in economics concerning qualitative choice models including choice of housing (choice between buying a certain good or a substitute), transportation (choice between individual or public transportation on the way to work) and labor market (choice of a worker whether to take a job offer or not). For example, Train (1986), investigates household vehicle demand. Some of his models estimate the probability of a household to own one, two or more automobiles, others estimate the probability of a household to choose one vehicle amongst a set of classes of different vehicles (foreign *vs* domestic, or choice between automobiles of different sizes are examples). Horowitz (1993) presents a model to estimate the probability of choice between automobile and transit for travel to work. van Soest (1992) fits a neoclassical structural model of labor supply (or choice between labor and leisure) of both spouses. Berridge (1993) models job security.

INTRODUCTION

Parametric models like the probit and the logit are very popular because they are computationally tractable and easy to interpret. They rely on the specification of a certain distribution (respectively the standard normal or the logistic) for the probabilities of choice and the homoscedasticity (that is equal variance for all individuals) but real situations can be more complex. That is, heterogeneity among the preferences of the decision makers is likely to be present or the probabilistic structure of the model may not follow exactly the probit or logit classic specification.

An alternative to parametric models relying on much less assumptions are semiparametric models like the so-called single index. Briefly, these models consist in an unknown transformation of a linear function with an unknown finite number of parameters while parametric models like the probit or logit consist in a known transformation of the same sort of linear function. That is, semiparametric models do not make assumptions about distributional properties of the data and as it will be shown in chapter 2 they allow for heterogeneity in data. They rely much more on the information given by the data.

Parametric models have attractive features not shared by semiparametric models. A very important one is that the first allow a richer interpretation of the problem. Usually they are easier to estimate. They make possible to derive related models (for instance by calculating derivatives). However, a misspecified parametric model can mislead the nature of the data and lead to wrong inferences. Consequently, one should safeguard from misspecified parametric models.

The aim of this dissertation is to develop a set of tools that allow to test the specification of a parametric binary choice model within a semiparametric approach. This amounts to compare the parametric model with the semiparametric rival. The test procedure of Horowitz and Härdle (1994) is a privileged tool to pursuit this aim. The properties of this test procedure on models with binary responses were not studied by the authors. This work carries out a carefully study of those properties. It also proposes improvements to the test procedure of Horowitz and Härdle (1994) which enhances significantly its performance.

All along the manuscript two data sets are used to illustrate the procedures under analysis. These are the data about the choice of transportation in the way to work of Horowitz (1993) and credit–scoring of Fahrmeir and Tutz (1994). A description of these data sets is included in the introduction of chapter 1.

The plan of the thesis is the following. Chapter 1 introduces the binary choice model. The most well known parametric specifications are discussed.

An overview of models for discrete responses other than binary is also presented. Chapter 2 is devoted to the semiparametric binary choice model. Some basic concepts about nonparametric regression estimation are also introduced. Chapter 3 makes a brief outline of several testing procedures that can be applied to the main problem motivating this work and are in some way a source of inspiration to the test of Horowitz and Härdle (1994). There, it is explained why this test is preferable. The chapter proceeds with a detailed analysis of the test procedure. Chapter 4 analyzes the performance of the Horowitz and Härdle (1994)'s test on binary choice models. Chapter 5 introduces an improvement to this test based on bootstrap while chapter 6 presents some analytical corrections to the bias and variance of the test statistic which ameliorate significantly its behavior in finite samples. Chapter 7 applies the techniques under study to test the adequacy of the logit fit in two real data sets concerning respectively unemployment after apprenticeship and credit–scoring.

with a clear decreasing increment rate from a certain level on of the differential in costs which is due to the parametric specification chosen.

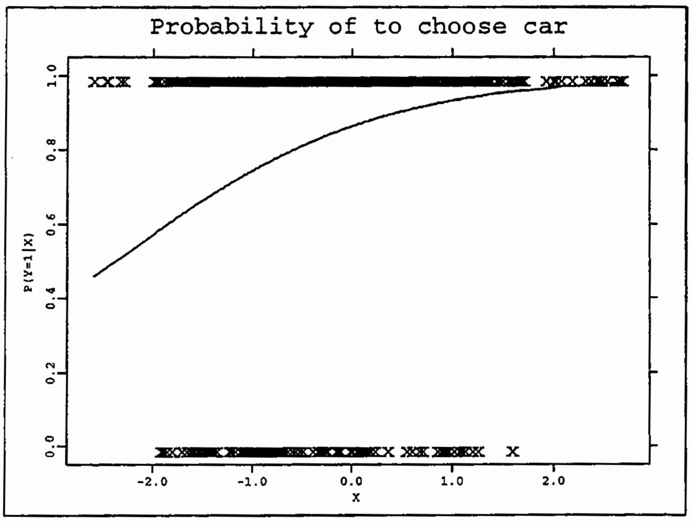

Figure 1.1: Logit fit of the probability of individual choice of car for travel to work as function of transit fare minus automobile cost.

Fahrmeir and Tutz (1994) introduce an example of credit-scoring. The aim is to estimate the probability of an individual that has borrowed a credit amount is to be considered a potential risk by not paying back the debt as agreed upon by contract. As before the probability is a function of a set of covariates which influence the credit-ability of the individual here considered as risk factors. These covariates include, among others, the following: running account with categories no, medium, and good; duration of credit in months; amount of credit; payment of previous credits with categories good and bad; intended use, with categories private and professional.

Figure 1.2 shows a fit of the mentioned probability as function of the amount of credit borrowed for the data in Fahrmeir and Tutz (1994). The fit was obtained by logistic regression. The probability of a client constitute a potential risk increases with the amount of credit borrowed and is always below 0.8.

Later on, within the chapter, the assumptions underneath the models just presented will be discussed, alternative models combining the effect of several covariates will be examined as well. The economic motivation and interpretation of the binary model are discussed together with the alternative

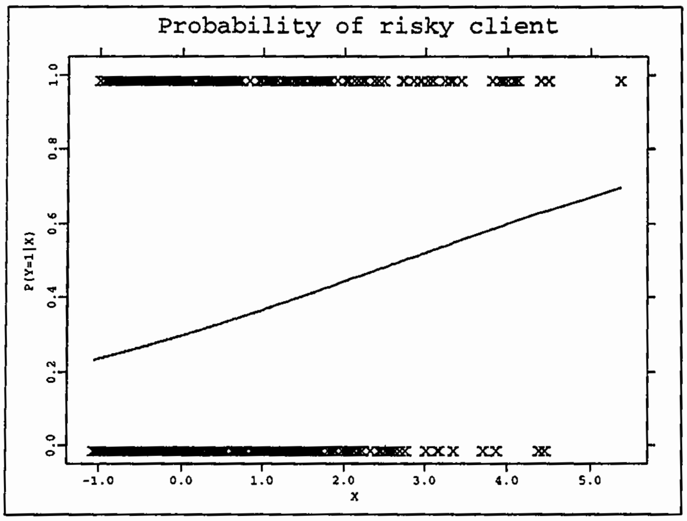

Figure 1.2: Logit fit of the probability of a client to be a potential risk as function of the amount of credit borrowed.

strictly econometric formulation. Parametric models like the probit, random-coefficients probit, logit and complementary log-log are analyzed.

The particular decision or attribute of the individual can be expressed by a random variable (the dependent variable) assuming only the values 0 or 1. When the individual has to choose in a set of more than two mutually exclusive alternatives or may verify one among more than two attributes the response is no longer binary. Models for categorical responses other than binary are much more complex. While this thesis concentrates on binary response models this chapter gives also a general overview of other categorical response models.

At the end of the chapter four different simulated data sets are introduced. They will be used during the work to illustrate the applicability of the methods in study.

1.2 The Utility Function Approach

The binary choice model can be derived from utility theory. This approach is very popular among econometricians due to its clear economic insight and can be found for instances in Amemiya (1981, 1983), Hausman and Wise (1978), Judge, Griffiths, Hill, Lütkepohl and Lee (1985), and Train (1986).

In this approach the binary choice model is derived based on certain assumptions that define the behavior of the individual decision makers. Suppose that an individual decision maker is confronted with the choice between two different (and mutually exclusive) alternatives or attributes. For sake of clearness lets identify one alternative by I and the other by II. Economic examples are, among others, the choice of an individual to travel to work by automobile (alternative I) or by transit (alternative II), the choice of an individual to buy a certain product (alternative I) or not (alternative II), the choice of an individual to participate in the labor market (alternative I) or not (alternative II), the choice of an enterprise to invest (alternative I) or not (alternative II). The individual choice is represented by a random variable, Y_i, that assumes the value $y_i = 1$ if one alternative is chosen, say alternative I, or $y_i = 0$ if the other is chosen, say alternative II.

Each individual has an indirect utility associated to each alternative represented by an utility function. The utilities depend on the particular characteristics of the individual decision maker and the specific attributes of the alternative. Relatively to the example introduced about the probability of mode choice to travel to work one may assume that each individual has an utility associated with travelling by automobile and another utility associated with travelling by transit. Both utility functions depend on exogenous variables characterizing the individual like the number of cars owned by the household and the respective mode of travel like the time and fare of automobile for the first and the time and fare of transit for the second.

To simplify the specification of the individual utilities a common procedure used in economics and econometrics is based on the "representative individual" approach. This approach postulates the existence of a "representative" or "average" individual who is supposed to have tastes equal to the average over all individual decision makers. Let us assume that the indirect utilities of the "representative" individual associated to alternative I, say V_1, and associated to alternative II, say V_0, depend on a set of exogenous variables z, representing the individual characteristics, and w_j with $j = 0, 1$ corresponding to the specific attributes of each alternative as faced by the individual according to

$$\begin{aligned} V_1 &= V(z, w_1, \gamma) \\ V_0 &= V(z, w_0, \gamma) \end{aligned} \tag{1.1}$$

where $V(\bullet)$ is a function known up to a vector of parameters γ with finite dimension. Usually one assumes that $V(\bullet)$ is a linear function.

It is further assumed that all individuals have a common structure in their utilities which is the "representative" individual utility structure, that is the function $V(\bullet)$. Moreover, the variation of the individual tastes from the "average" is captured in the utility function only by a random component not observable.

Define the indirect utility of the ith individual associated to alternative I for which $y_i = 1$ as U_{1i} and the indirect utility associated to alternative II corresponding to $y_i = 0$ as U_{0i}. The utilities of a given individual of the population, say individual i, have two components. One, depends only on factors observable by the econometrician and has a common structure for all individuals equal to the utility of the "representative" individual. This part is referred to as the representative utility and expresses the average behavior of the population. The other, contains all influences that are unknown and gives the deviation of the individual tastes from the average behavior. This part is non-observable and random. Therefore, individual utilities are stochastic.

According to the above reasoning the utilities of individual i can be expressed as

$$\begin{aligned} U_{1i} &= V(z_i, w_{1i}, \gamma) + \epsilon_{1i} \\ U_{0i} &= V(z_i, w_{0i}, \gamma) + \epsilon_{0i} \end{aligned} \tag{1.2}$$

where $V(\bullet)$ are the representative utilities and ϵ_{ji}, $j = 1, 0$, are random variables representing the stochastic part of the utilities which reflect the random tastes of the individual.

The aim is to fit the probability that the ith individual chooses alternative I, that is $P(Y_i = 1)$. The individual i will choose alternative I if the indirect utility associated to it is greater than the indirect utility associated to the other alternative. The aimed probability becomes $P(U_{1i} > U_{0i})$ or according to (1.2)

$$P\{\epsilon_{1i} - \epsilon_{0i} > V(z_i, w_{0i}, \gamma) - V(z_i, w_{1i}, \gamma)\} \tag{1.3}$$

This probability depends on the distribution of the individual utilities, more precisely, on the distribution assumed for the deviation of the random tastes, $\epsilon_{1i} - \epsilon_{0i}$. Then

$$P(Y_i = 1) = 1 - G\{V(z_i, w_{0i}, \gamma) - V(z_i, w_{1i}, \gamma)\} \tag{1.4}$$

with $G(\bullet)$ the distribution function of $\epsilon_{1i} - \epsilon_{0i}$. To conclude, the specification of the binary choice model depends on the particular function assumed for the representative utilities, $V(\bullet)$, and the particular distribution assumed for the random utilities or more precisely the distribution $G(\bullet)$.

1.3 The Latent Variable Model

The binary choice model can be formulated by a latent variable model as in Maddala (1983). Suppose that underlying the choice variable Y_i which assumes the values 1 or 0 there is a real-valued random variable Y_i^* known as the latent variable, such that

$$\begin{aligned} y_i &= 1 \quad if \;\; y_i^* > 0 \\ y_i &= 0 \quad otherwise \end{aligned} \tag{1.5}$$

The usual approach assumes that the latent variable has a linear behavior defined by the relationship

$$Y_i^* = X_i^T \beta + u_i$$

where X_i is a vector of observable exogenous random variables taking values in $\mathbb{R}^k$ which express the individual and the alternative characteristics, and u_i is a real-valued non-observable random variable whose stochastic structure will be examined later. From now on we will assume that Y_i^* has the linear behavior given in last equation.

The probability that individual i chooses alternative I given the values observed for the exogenous variables is,

$$P(Y_i = 1|X_i = x_i) = P(Y_i^* > 0|X_i = x_i) = P(u_i > -x_i^T \beta) \tag{1.6}$$

In practice the latent variable Y_i^* is non-observable. Furthermore, to estimate the probability model (1.6) is not necessary to know the particular value assumed by the latent variable. In the following it will be shown how the latent variable is related to the individual utilities which are very difficult to observe in real situations.

The latent variable model and the formulation based on the utility theory are closely related. Assuming that the representative utility in (1.2) is linear then the deviation in the representative utilities of individual i becomes

$$V(z_i, w_{1i}, \gamma) - V(z_i, w_{0i}, \gamma) = z_i^T(\gamma_1 - \gamma_0) + (w_{1i} - w_{0i})^T \bar{\gamma}$$

where γ_0 are the elements of γ that are coefficients of the variables in z_i in utility U_{0i}, γ_1 are the elements of γ that are coefficients of the variables in z_i in utility U_{1i} and $\bar{\gamma}$ are the remainder elements of γ which are the coefficients of variables in w_{ji}, $j = 0, 1$. Consider that $\gamma_1 - \gamma_0$ and $\bar{\gamma}$ constitute the vector of parameters β, that the vector of exogenous x_i is formed by variables z_i

(like the number of cars owned by the household in the running example) and $(w_{1i} - w_{0i})$ (like transit travel time minus automobile travel time and transit travel fare minus automobile travel fare in the mentioned example) and finally put $\epsilon_{1i} - \epsilon_{0i} = u_i$, with x_i, u_i and β as in model (1.6). With this change of variables the deviation of representative utilities presented above becomes just $x_i^T\beta$ and consequently the probability that $U_{1i} > U_{0i}$ in equation (1.3) reduces to $P(u_i > -x_i^T\beta)$ which is also the probability $P(Y_i^* > 0|X_i = x_i)$ in equation (1.6). Therefore the distribution $G(\bullet)$ of $\epsilon_{1i} - \epsilon_{0i}$ conditional on $X_i = x_i$ is the distribution function of u_i conditional on the same variable. From now on each time the distribution of u_i or $\epsilon_{1i} - \epsilon_{0i}$ is referred to the distribution conditional on $X_i = x_i$ of those random variables is meant.

Note that the linear function $x_i^T\beta$ can be generalized in order to include some interaction terms or known transformations of explanatory variables. These interactions and transformed variables are treated merely as new exogenous variables enlarging the vector of explanatories x_i. Also, except if stated otherwise, it is assumed that x_i has as first element 1 in order to include an intercept term.

Very often the probability of choosing alternative I (1.4) is defined with respect to the distribution function of $\epsilon_{0i} - \epsilon_{1i}$ or equivalently to the distribution of $-u_i$ normalized by a certain "convenient" variance resulting in the function $F(\bullet)$.

Considering a sample of n individuals and assuming that their random utilities are all identically distributed the binary choice model is defined by

$$P(Y_i = 1|X_i = x_i) = F\left(\frac{x_i^T\beta}{\sigma}\right) \quad i = 1, \cdots, n$$

with σ^2 the variance of u_i weighted by the particular variance that scales $F(\bullet)$.

In practice we are not able to identify σ^2 which is absorbed by the coefficient values for the explanatories. Consequently, from now on, to ease the notation in the homoscedastic case where utilities are identically distributed, the coefficients β are supposed to be normalized by σ. This amounts to consider that the probability of choice is given simply by the equation

$$P(Y_i = 1|X_i = x_i) = F(x_i^T\beta) \quad i = 1, \cdots, n \tag{1.7}$$

The analysis of equation (1.7) suggests some comments. First, the relation with equation (1.4) reveals that the probability of choice of an alternative depends only on the difference in the individual utilities associated to each alternative and not on their absolute value. Consequently the choice of explanatory variables is restricted to the choice of variables that describe the ordered or sequential. Because the focus of this thesis is on binary response models the survey in this section has not the aim to provide a deep and detailed study but to give a general overview on models for multinomial responses.

As before, the individual will choose the alternative that maximizes his utility and utilities have a deterministic part giving the representative utility or the average behavior and a unobserved and random part which is the deviation in individual tastes from the average behavior.

Suppose that the individual i faces m different and mutually exclusive alternatives. Let $P_{ji} = P(Y_i = j|X_i = x_i), j = 1, \cdots, m$, be the probability that individual i chooses alternative j and assume that this probability is a function of linear indexes according to $P_{ji} = F_j(x_{1i}^T\beta, \cdots, x_{mi}^T\beta)$, with $x_{ji}, j = 1, \cdots, m$, a vector with explanatory variables with the same dimension as β. One can assume that the observations for the responses, y_i, are arising from a multinomial distribution defined on m different categories where category j has probability P_{ji} for individual i.

Parametric models assume a known form for $F_j(\bullet)$ and the estimation of the probabilities of interest is reduced to the estimation of the parameters β by the maximum likelihood method. To obtain the maximum likelihood estimator it is useful to define the following set of dummy variables

$$\begin{aligned} y_{ji} &= 1 \quad \textit{if} \quad y_i = j \\ y_{ji} &= 0 \quad \textit{otherwise} \quad j = 1, \cdots, m \quad i = 1, \cdots, n \end{aligned}$$

Then the log likelihood function may be written as

$$l = \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ji} \log P_{ji} \tag{1.17}$$

Maximization of the log likelihood (1.17) is much more complicate than the maximization of the log likelihood of the binary model. The complexity depends also on the particular parametric model considered for $F_j(\bullet)$ and on the structure of the alternatives or attributes (whether they are unordered, ordered or sequential). Details can be found for example in Amemyia (1983).

1.6.1 Unordered Multinomial Models

An example of unordered alternatives can be found in Hausman and Wise (1978). The problem they study is the individual choice of mode to travel to work. Individuals face three alternatives: driving the own car, sharing rides, and riding a bus. There is no order relation between these alternatives.

The most common model in this problem is known as the multinomial logit model. The multinomial logit is often used in practice because the calculations necessary to obtain the maximum likelihood estimates are simpler than in other common parametric models like the multinomial probit. The model verifies,

$$
\begin{aligned}
P_{ji} &= \frac{\exp(x_{ji}^T\beta)}{1+\sum_{j=1}^{m-1}\exp(x_{ji}^T\beta)} \qquad j = 1,\cdots,m-1 \\
P_{mi} &= \frac{1}{1+\sum_{j=1}^{m-1}\exp(x_{ji}^T\beta)}
\end{aligned} \tag{1.18}
$$

Amemyia (1983) shows how the multinomial logit can be derived from utility maximization.

The multinomial logit assumes that random utilities related to each alternative are independent. This assumption implies that alternatives are dissimilar. Suppose that an individual in choosing the transportation mode to work is faced with the alternatives own car, bus, and train. The probability of driving the car is $P_1 = P(U_1 > U_2, U_1 > U_3)$ with U_1, U_2, U_3 the utilities associated to driving the own car, riding by bus and riding by train. The multinomial logit assumes that the events $(U_1 > U_2)$ and $(U_1 > U_3)$ are independent meaning that riding by bus has no common characteristic for the individual with riding by train. However those alternatives are not completely dissimilar given that both represent public transportation. Consequently, the multinomial logit is underestimating P_1 because it ignores that if the individual prefers the car to the bus and $U_1 > U_2$ makes $U_1 > U_3$ more likely.

McFadden has called this characteristic of the multinomial logit the "independence from irrelevant alternatives" (IIA) (McFadden, 1981). When some alternatives are similar a model should be used that to some extent does not verify this property. This is the case for the generalized extreme-value (GEV) model and the conditional probit (already referred to for the binary problem). The first model is derived assuming that random utilities have a GEV distribution and has as particular formulations the nested logit model and the higher-level nested logit model described in Amemyia (1983). The second assumes that random utilities have a join normal distribution with a certain covariance matrix allowing for correlation among different alternatives and is described in Hausman and Wise (1978).

	coefficients	st. errors
Intercept	-1.2216	0.3022
Number of cars	2.3081	0.2243
Out-of-vehicle travel time	0.0622	0.0173
In-vehicle travel time	0.0092	0.0095
Travel cost	0.0169	0.0021

***Table 1.1:** Results of the logit fit on the problem of mode choice for travel.*

by the household, transit travel time relatively to automobile and transit fare relatively to automobile costs although in-vehicle travel time may not be statistically significant to explain the probability of choice. Figure 1.4 shows the logit fit for this example. Now the probability curve is plotted against the fitted index.

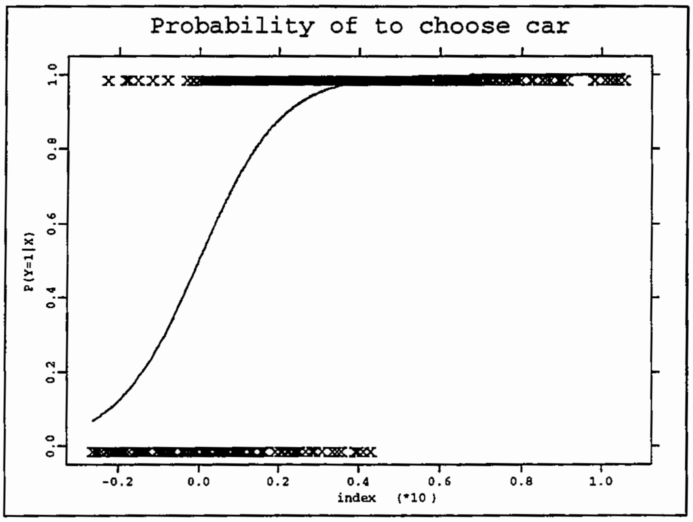

***Figure 1.4:** Logit fit of the probability of individual choice of car for travel to work.*

Table 1.2 shows the estimates obtained for the logit fit of the credit-scoring data while the fit is plotted in Figure 1.5. The probability of a client to be considered a potential risk is a function of the duration of credit in months, payment of previous credits with values 0 for good and 1 for bad, amount of

	coefficients	st. errors
Intercept	-1.6232	0.3428
Duration of credit	0.0253	0.0078
Payment of previous credits	1.2900	0.2359
Amount of credit	0.0001	0.0000
Monthly payment	0.2080	0.0734
Age	-0.0215	0.0069

***Table 1.2:** Results of the logit fit on the problem of credit-scoring.*

credit, percentage of monthly payment in the individual income and age of the client. The probability of a client to be a potential risk grows with the duration of credit, a bad payment of previous credits, the amount borrowed and the weight of the monthly payment on individual income. On the other side the probability decreases with the age of the client.

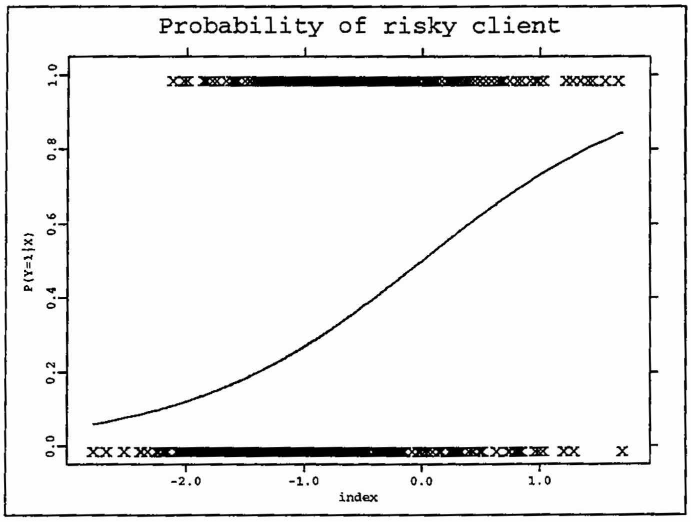

***Figure 1.5:** Logit fit of the probability of a client to be a potential risk.*

1.8 An Example

This section introduces four simulated data sets which differ in the link function considered. These data sets will be used in the remainder of this thesis to illustrate the reasoning of the methods to be presented. They are described in the following.

Four parametric models were considered for the conditional probability of choice of alternative I, $F(x_i^T\beta)$ according to,

$$
\begin{aligned}
F(x_i^T\beta) &= \{1+\exp(-x_i^T\beta)\}^{-1} && (1.21)\\
F(x_i^T\beta) &= 1-\exp\{-\exp(x_i^T\beta)\} && (1.22)\\
F(x_i^T\beta) &= \{1+\exp(-x_i^T\beta)\}^{-1} + \frac{-x_i^T\beta}{1.5}\times\frac{1.25}{1.5}\times\varphi\left(\frac{x_i^T\beta}{1.5}\right), && (1.23)\\
F(x_i^T\beta) &= \left[1+\exp\left\{\frac{-x_i^T\beta}{s(x_i^T\beta)}\right\}\right]^{-1} && (1.24)\\
s(x_i^T\beta) &= (1+|x_i^T\beta|)^{1/2}
\end{aligned}
$$

with $\varphi(\bullet)$ the standard normal density.

Models (1.21) and (1.22) are classic parametric models, the logit and the CLL. Model (1.23) is a logit model perturbed by a bump with height given by 1.25 and width equal to 1.5. The value 1.25 was chosen such that the conditional expectation of Y_i will never decrease when $x_i^T\beta$ increases. Finally, model (1.24) is a logit model with heteroscedasticity where the heteroscedasticity is given by the function $s(x_i^T\beta)$.

These models were chosen because they may be considered typical to characterize some frequent situations that may arrive in problems with binary responses.

The logit model is the most used for problems with no presence of heterogeneity in data which are well depicted by a symmetric link while the CLL is a popular alternative model when a non-symmetric link is required.

The logit with bump link in (1.23) is a deviation from the logit link which make the conditional probability function to be flatter on a neighborhood of the index function centered at zero as can be seen in the lower left plot of Figure 1.6. In that region the link has an increasing rate almost equal to zero. Mind that the index function translates the difference in the utility associated to each alternative. In general it may be interpreted as the difference in the score assigned by the individual to each alternative. Consequently, when the index is around zero it means that the score allocated by the individual to each alternative is very similar. It is natural to think that individuals have some difficulties to decide which alternative to choose in those cases. This behavior can be translated by a weak increase of the cumulated conditional probability of choice, or by the flatness of the link function, in the region where the index assumes values around zero and can be represented by the logit with a bump link in (1.23). On the other side, this model can be viewed as giving merely an alternative behavior of individuals relatively to the logit where individuals prefer more strongly alternative one under an unfavourable score (when the index assumes negative values) and prefer it less under a favourable score of this alternative relatively to the other (when the index assumes positive values) than those individuals behaving according to the logit model.

The logit with heteroscedasticity represents a deviation from the logit model that incorporates heterogeneity among individuals by considering an heteroscedastic latent variable or heteroscedastic stochastic utilities. The presence of heterogeneity among individuals may be relevant in practical situations. See as example the application on mode-choice of travelling in the way to work of Horowitz (1993).

For all experiments the index function was assumed to be $1 - x_{i1} + 2x_{i2}$ and the number of observations was set to 500. The regressors were generated independently from a standard normal. For sake of comparison the same data set for the regressors was used in all experiments to generate the response from a Bernoulli distribution with probability of success given respectively by models (1.21) to (1.24). Thus, individual i has explanatory variables with the same magnitude in all data sets, with $i = 1, \cdots, 500$. On the othr hand, the endogenous variable was generated using the same random seed in all experiments.

Figure 1.6 shows the four models introduced. The logit with a bump (lower left), the logit with heteroscedasticity (upper right) and CLL (lower right) are plotted together with the logit model.

In practice the true link function is unknown and a common behavior is to assume that data come from a logit specification. Assuming a logit link results in a misspecification when estimating data sets generated by models (1.22), (1.23), and (1.24).

Figure 1.7 shows the shapes of the introduced models with the parametric logit fit obtained in each data set respectively. The logit fit was plotted against the estimated index, $x_i^T\hat{\beta}$, while the true models were plotted against the true index $x_i^T\beta$ (which is the same in all experiments according to what was said before). The upper left shows data generated from the logit estimated assuming a logit. The fit and the true model are coincident. All the other plots in the

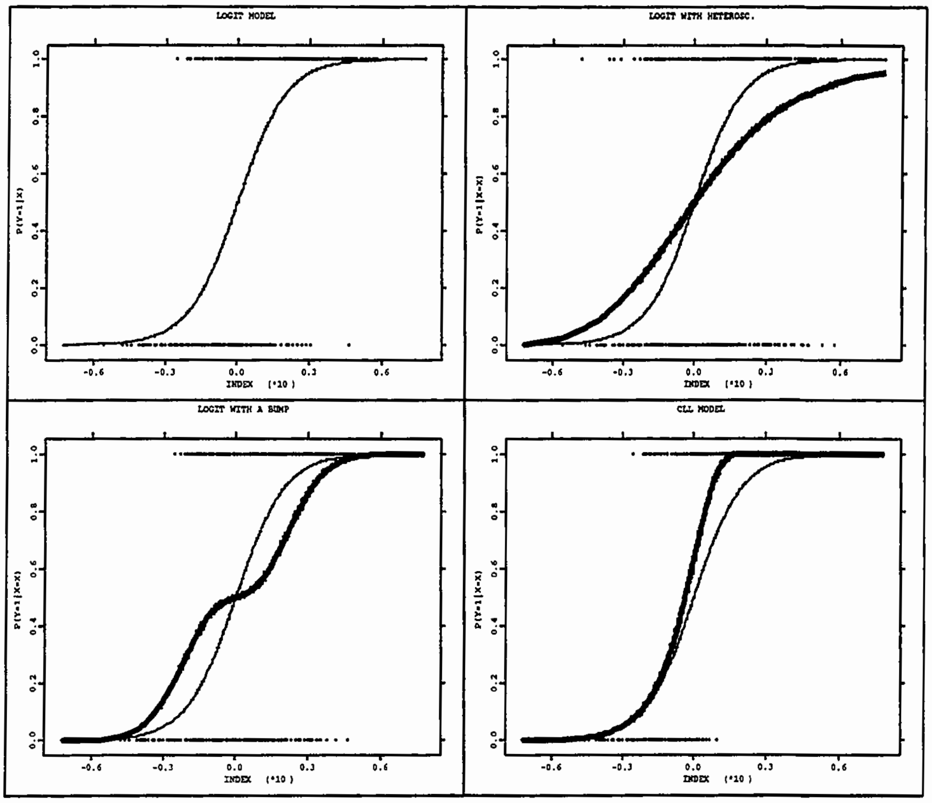

Figure 1.6: Upper left – the logit model, lower left – the logit with a bump (thick line) and the logit, upper right – the logit with heteroscedasticity (thick line) and the logit, lower right – the complementary log–log and the logit. The values for the response generated by the model in the plot are identified by crosses.

figure show misspecified fits.

A feature to remark in Figure 1.7 is that the length of the fitted index for the misspecified models in the upper right and lower left, identified in the figure by the range of the support of the logit fit, shows a tendency to underestimate the range of the true index. This happens because in fact the variance of the latent is different in all models given that β is the same for all but each link is defined for a different scale. When estimating the data with a logit link the estimates of the standard deviation of the latent will automatically scale the estimates of β (weighted by the the standard deviation of the logistic distribution). Therefore, the scale of the estimates for β or equivalently the scale of the fitted index, has

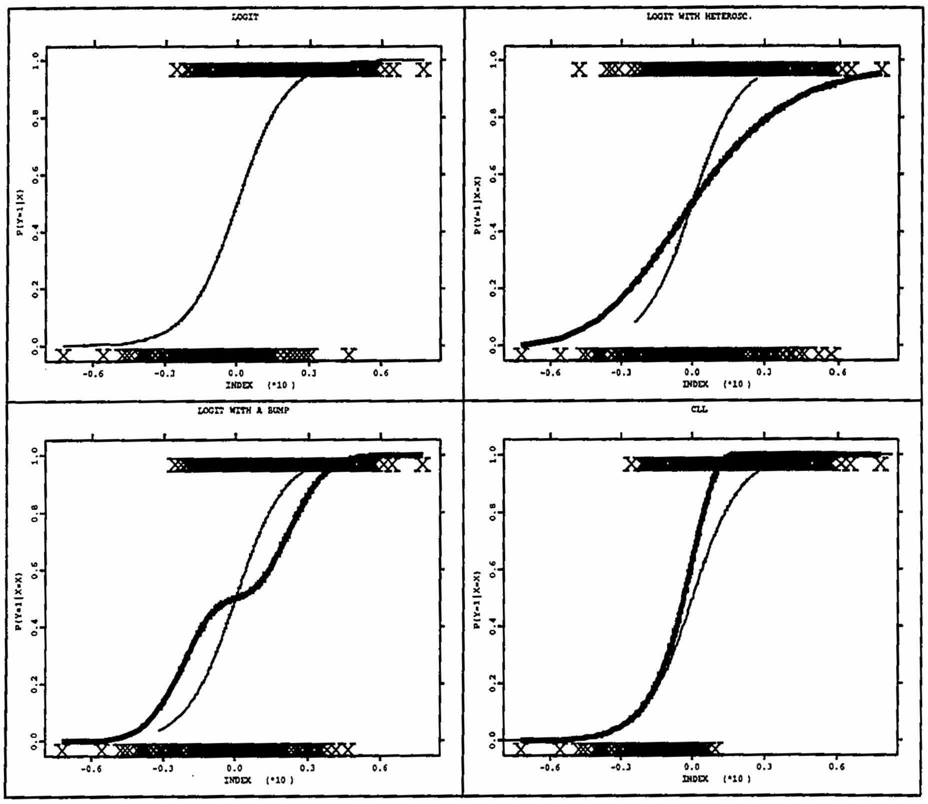

Figure 1.7: Upper left – the logit model, and logit fit, lower left – the logit with a bump (thick line) and parametric logit fit, upper right – the logit with heteroscedasticity (thick line) and parametric logit fit, lower right – the complementary log–log and parametric logit fit. The values for the respective response in each plot are identified by crosses.

to be different in each specification.

Plots in Figure 1.7 may lead to an erroneous judgment of the kind of the deviation the logit fit has from the true model. To compare this deviation one has to normalize the fitted index or alternatively the true index in order that β and $\hat{\beta}$ are in the same scale by making the necessary adjustments with the standard deviations of the true and the logit links. Figure 1.8 shows the plots with the true index normalized by dividing by the standard deviation of the respective true link and multiplying by the standard deviation of the logit link. The CLL was also adjusted in location.

The lower left of Figure 1.8 shows that the bump in the shape of the true

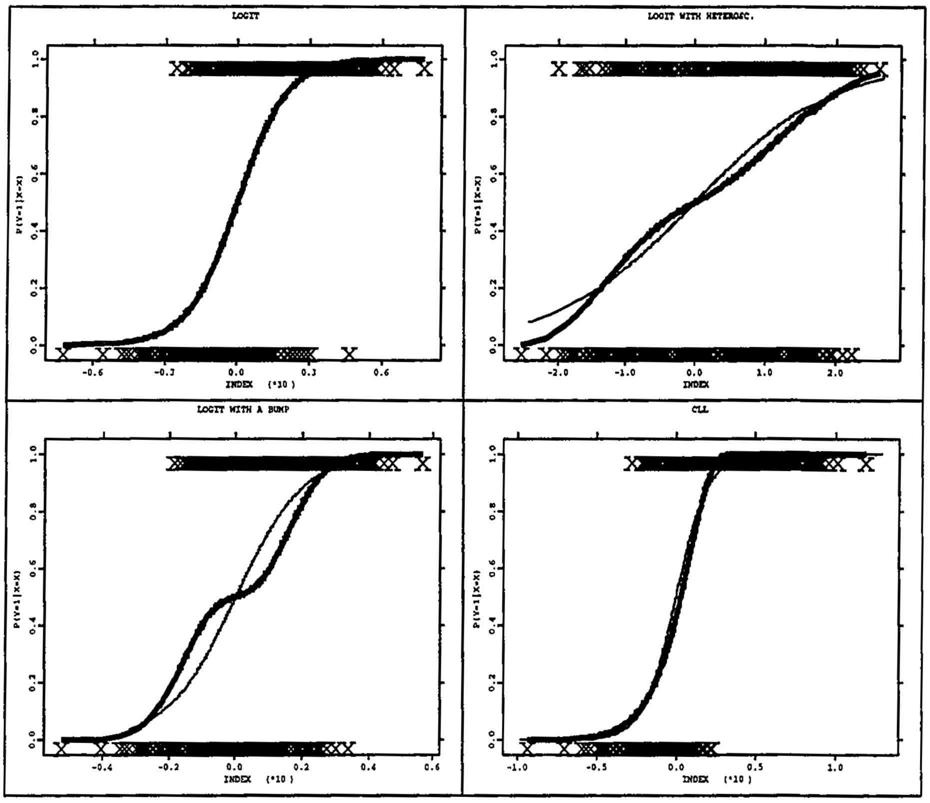

Figure 1.8: Upper left – the logit model, and logit fit, lower left – the logit with a bump (thick line) and parametric logit fit, upper right – the logit with heteroscedasticity (thick line) and parametric logit fit, lower right – the complementary log–log and parametric logit fit. The values for the respective response in each plot are identified by crosses. Index normalized.

link is not captured by the ill-specified parametric fit. The upper right shows the logit with heteroscedasticity. Note that plotting the probabilities against the index divided by the heteroscedastic variance function modifies the shape of the curve. Now the curve shows a bump which is not captured by the logit fit. However the deviation is not so great as in the case before. In the lower right the CLL model with misspecified logit fit is given. Here, the misspecification of the fit is not explicit given that the true CLL model is very close to the logit fit.

The plot in the lower right shows that the location of the CLL is very well

	intercept	X1	X2
logit	0.7870	-1	1.4601
	(0.1093)	(0.12235)	(0.1413)
logit with	0.4933	-1	1.3317
bump	(0.1557)	(0.1669)	(0.1733)
logit with	0.5162	-1	1.1725
heterosc.	(0.1704)	(0.1809)	(0.1821)
CLL	1.2772	-1	1.5647
	(0.1321)	(0.1220)	(0.1571)

Table 1.3:* *Results of the logit fit on the simulated data. Standard errors inside brackets.

estimated by the logit fit. Mind that the location is absorbed into the estimate of the constant term. The logit with a bump and logit with heteroscedasticity links are referred to the same location as the logit. Thus, is natural that the logit fit of data from those models has no location inaccuracy as the figure shows.

All the plots show that the spread of the fitted index is very close to the spread of the normalized true index which induces that the misspecified logit fit is able to assess correctly the variance of the latent variable (though it is not identifiable in practice). Mind that in the plots in the lower left the true index was normalized by multiplying β by a constant while in the lower right it was normalized by adding a constant and multiplying by another. Therefore one can conclude that in those examples the slope coefficients in β are well approximated up to a proportional constant by the respective elements of $\hat{\beta}$. This subject will be examined later on in the next chapter. For the heteroscedasticity case in the upper right the same conclusion holds if one considers that the heteroscedasticity is reflected in the distortion of the logit link as is shown in the plot. More precisely, the function $s(\bullet)$ is absorbed into the link which then is not anymore a logit as in the right upper plot of Figure 1.6.

Table 1.3 shows the logit estimates for the coefficients β in the four data sets introduced. The estimates are normalized from being divided by the absolute value of the coefficient estimate of the first regressor in order to be on the same scale for the different data sets. The logit fit was implemented in XploRe 3 using the GLM module. The estimates are not much different except the estimate of the intercept of the CLL model which is clear larger than the others.

However, even if the misspecified logit fit gives accurate estimates of the slope coefficients in β up to a constant the true probability curves are deviating from the logit fit because of not taking into account the existence of bumps in the link for the logit with a bump and logit with heteroscedasticity models. When data are generated by the CLL model it is hard to distinguish the difference between the true link and the misspecified parametric estimate by a merely inspection of their plot. However, that distance may be assessed using another tool like a test statistic.

This work aims to study and to improve tools that allow to detect misspecification on the link function of a binary choice model by evaluating if the deviation between a parametric estimate of the link and the true function is statistically significant in the sense that is not due only to sample randomness.

Chapter 2

The Semiparametric Binary Choice Model

2.1 Introduction

In chapter 1 some parametric binary choice models were introduced. First, models that do not allow for the existence of heterogeneity on tastes of the individuals were focused. These are the logit, probit, and CLL. While results are almost the same whether probit or logit are used this is not the case with the CLL. Second, heterogeneity among individuals was introduced with the RCP model. Here heterogeneity appears as heteroscedasticity verifying a very particular parametric form namely the variances of the latent variables Y_i^* are given by $x_i^T \Sigma x_i$ with Σ a $k \times k$ positive definite matrix.

All the parametric models are derived based on distributional assumptions of the latent variable Y_i^* (or equivalently of the random utilities). These assumptions are somehow restrictive since they imply certain behavior of the individuals and may induce misspecification of the parametric model. In this case, maximum likelihood estimates may be inconsistent or inefficient engendering predictions about the individual choice that can be entirely wrong.

To be robust to the kind of misspecification mentioned above one can define a model where no assumptions (or very few) are made about the distribution of the latent variable. In this chapter a model will be presented that fulfills this aim. It widens the parametric assumptions about the individuals behavior in order to be more general and more flexible than the parametric models presented within a semiparametric approach. It also includes the parametric models as particular cases. This model is recognized in the literature as single index model (SIM) and can be applied to a wide class of problems from which binary responses are just a particular case.

The specification of the semiparametric model and its estimation will be addressed within the chapter. The estimation methods that are $\sqrt{n}$-consistent are privileged for reasons that will become more clear in the subsequent chapters. The construction of confidence bands based on the semiparametric fit will be also discussed and introduced as a first tool to compare a parametric model with the semiparametric rival.

2.2 The model

The semiparametric model can be viewed as a generalization of the parametric model. The purpose of the semiparametric approach is to widen the assumptions regarding the link function $F(\bullet)$ while avoiding the curse of dimensionality that is hampering fully nonparametric techniques when applied to high-dimensional data. The semiparametric model overcomes the curse of dimensionality by aggregating the multidimensional variable X_i into the single (parametric) index $x_i^T\beta$, while maintaining the nonparametric assumption that the specification of the link in equation (1.9) is unknown. A SIM formulation of the binary choice model can be seen for example in Stoker (1992).

The single index model for binary responses takes on the following form

$$E(Y_i|X_i = x_i) = P(Y_i = 1|X_i = x_i) = F(x_i^T\beta) \quad i = 1, \cdots, n \tag{2.1}$$

with $F(\bullet)$ an unknown ("smooth") function with range contained in $[0,1]$ which, as in (1.9), is not necessarily a distribution function.

The SIM may incorporate heterogeneity in tastes across individuals if it is manifested as heteroscedasticity. The heteroscedasticity is of unknown form and has to depend on the index function. To make more clear note that model (2.1) can be written

$$E(Y_i|X_i = x_i) = g\left\{\frac{x_i^T\beta}{\sigma(x_i^T\beta)}\right\} \quad i = 1, \cdots, n$$

with $g(\bullet)$ assuming values on $[0,1]$, $\sigma(\bullet)$ always positive, and both unknown functions. Here, if $\sigma(\bullet)$ is not equal to a constant it can be seen as the variance function of a heteroscedastic latent variable. In practice the function $\sigma(\bullet)$ is not identifiable and it is absorbed into the unknown link resulting in $F(\bullet)$. The main restriction of the SIM model (2.1) is the linearity of the utility functions or the linearity of the latent variable Y_i^*.

In the SIM the intercept is not identifiable and it is subsumed in the link function. Note that the mean of the random part of the latent variable, more precisely the mean of the variable $-u_i$ with u_i as in (1.5), is unknown. When the link $F(\bullet)$ is parametric it has a known fixed location which may not be the true mean of $-u_i$. The difference between the fixed location in $F(\bullet)$ and the true mean of $-u_i$ is given by the intercept of the model. Thus, in parametric models the intercept is normalized not only by the scale imposed to the link but also by its fixed location. For example, probit and logit models have fixed location equal to zero therefore the intercept gives the mean of $-u_i$ for a given scale. Because in the SIM $F(\bullet)$ is free there is not an automatically fixed location allowing to identify the intercept which is instead absorbed in $F(\bullet)$. The same will happen to the scale. That is, because $F(\bullet)$ is unknown it can not impose a natural scale and location for the index like in parametric models. To conclude, for parametric models the link $F(\bullet)$ is defined for a given location and scale which fixes a natural normalization for the coefficients β, that is the scale and the intercept value. In the SIM the procedure is inverse. Because the link is unknown it is necessary to define a convenient normalization for β exogenously which fixes naturally the scale and location of the link $F(\bullet)$. This means that scale and location will be absorbed into the link.

One strategy to normalize the vector β consists on fixing the intercept equal to zero and one of the other component coefficients equal to 1. This will be the normalization used in this work. Therefore, the vector of coefficients β (without intercept) is uniquely identified up to a multiplicative constant. From now on it will be considered that the vector of coefficients β in the SIM has no intercept.

2.3 Semiparametric estimation of the SIM

Estimation of the semiparametric model (2.1) proceeds in two steps. First, the coefficient vector β has to be estimated. Let us call this estimate $\hat{\beta}$ and $x_i^T\hat{\beta}$ the estimated index. The second step estimates the link function by smoothing the data y_i on the projected index $x_i^T\hat{\beta}$. The smoother used in this work is the

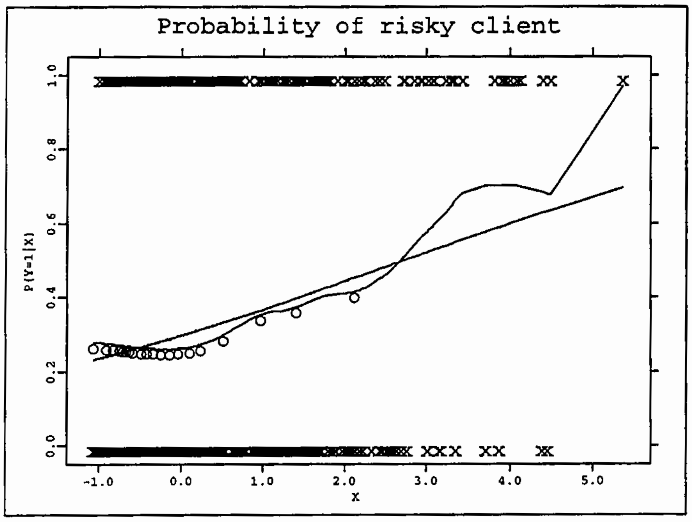

Figure 2.2: Probability curve of a client being a non potential risk as function of the amount of credit borrowed: logit fit and kernel regression.

2.3.1 WADE Estimator

This method is a modification of average derivative estimation (ADE) (Härdle and Stoker, 1989). ADE is motivated by the following property of single index models of the form (2.1)

$$E\{\nabla F(x_i)\} = E\left(\frac{dF}{dx_i^T\beta}\right)\beta = \gamma\beta$$

where the expectation is taken with respect to the distribution of X_i and $\nabla F(x_i) = \partial F/\partial x_i$.

Assuming that $F(\bullet)$ is a.e. first differentiable in X_i, β can be estimated up to a constant by estimating the mean of the gradient vector $\nabla F(x_i)$. On the other hand, if $\nabla F(x_i)$ is proportional to β then any weighted average of the derivatives $\nabla F(x_i)$ will also be proportional to β. Let $\omega(x_i)$ be a weighting function. Then

$$E\{\omega(x_i)\nabla F(x_i)\} = E\left\{\omega(x_i)\frac{dF}{dx_i^T\beta}\right\}\beta = \gamma_\omega\beta$$

This equation motivates the WADE procedure. The aim is to choose a convenient weight function in order to make the estimation of the mean of the weighted gradient vector easier than the estimation of the mean of its unweighted counterpart. Powell et al. (1989) show that this is accomplished if the weight function is the density function of X_i.

Suppose that X_i is continuously distributed with density $p(x_i)$ which is also differentiable. Put $\omega(x_i) = p(x_i)$. Under some suitable regularity conditions (see Powell et al., 1989) integration by parts allows to write

$$E\{p(x_i)\nabla F(x_i)\} = -2E\{Y_i \nabla p(x_i)\}. \tag{2.5}$$

with $\nabla p(x_i)$ the gradient vector of $p(x_i)$. Therefore, estimating β up to a constant amounts to estimate the derivative of the density of X_i.

Given a sample of n individuals the WADE estimator (with weight $p(x_i)$) for a constant times β is given by

$$\hat{d} = \frac{-2}{n} \sum_{i=1}^{n} y_i \widehat{\nabla p}(x_i). \tag{2.6}$$

The estimate of the gradient of the density of X_i is obtained by kernel smoothing. For theoretical reasons the estimate at point x_i omits the ith observation in the smoothing process according to

$$\widehat{\nabla p}(x_i) = \frac{1}{n-1} \sum_{j \neq i}^{n} \left(\frac{1}{h}\right)^{k+1} K'\left(\frac{x_i - x_j}{h}\right)$$

where k is the number of explanatory variables in X_i.

Powell et al. (1989) show that $\hat{d}$ is asymptotically normal distributed with asymptotic covariance matrix consistently estimated by

$$\hat{\Sigma}_d = \frac{4}{n} \sum_{i}^{n} \hat{r}(x_i)\hat{r}(x_i)^T - 4\hat{d}\hat{d}^T \tag{2.7}$$

where

$$\hat{r}(x_i) = \frac{1}{n-1} \sum_{j \neq i}^{n} \left(\frac{1}{h}\right)^{k+1} K'\left(\frac{x_i - x_j}{h}\right)(y_i - y_j)$$

Note that WADE can be applied only to "continuous" explanatory variables. That means, categorical regressors or dummy variables cannot be included in the model. For some problems this can be an undesirable restriction.

2.3.2 Maximum Quasi-Likelihood Estimator

Klein and Spady (1993) propose as estimate of β the value that maximizes the log likelihood of the binary choice model (1.14) when $F(x_i^T\beta)$ is substituted by $\widehat{F}_{hi}(x_i^T\beta)$ a nonparametric kernel regression with bandwidth h, calculated excluding observation i and usually known as leave-one-out (LOO) kernel regression. The LOO estimator is defined by,

$$\widehat{F}_{hi}(v) = \frac{\sum\limits_{j \neq i}^{n} K\left\{(v - x_j^T\beta)/h\right\} y_j}{\sum\limits_{j \neq i}^{n} K\left\{(v - x_j^T\beta)/h\right\}} \tag{2.8}$$

where all the variables have the same meaning as before.

The introduction of the above estimate in the likelihood function results in a so-called quasi-likelihood. For theoretical purposes the authors introduce a trimming in the quasi-likelihood to eliminate observations with imprecise estimators of $F(\bullet)$. In practice the trimming has revealed to be not important and can be omitted.

The maximization of the quasi-likelihood is performed iteratively in the same way as any nonlinear function. Bonneu, Delecroix and Malin (1993) advise to consider for the bandwidth h in each iteration the value $\hat{s}(x_i^T\bar{\beta})n^{-1/5}$ where $\hat{s}(x_i^T\bar{\beta})$ is the sample standard deviation of the index $x_i^T\bar{\beta}$ $i = 1, \cdots, n$ with $\bar{\beta}$ the actual estimate of β in that iteration.

Klein and Spady (1993) prove the asymptotic normality of the quasi-likelihood estimator $\hat{\beta}$. They show that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

with

$$\Sigma = E\left\{\left(\frac{\partial F}{\partial \beta}\right)\left(\frac{\partial F}{\partial \beta}\right)^T \frac{1}{F(1-F)}\right\}^{-1} \tag{2.9}$$

To estimate the covariance matrix Σ it suffices to plug-in in (2.9) the standard estimates of the unknowns, respectively β and F and take the sample mean. Because the information equality between the negative of the Hessian matrix and the expected outer product gradient still holds in this problem the covariance above can be also estimated with White's (1982) estimator. Note that for inference purposes the quasi-likelihood may be treated like an usual likelihood function. Therefore, the estimates for the variance of the coefficients returned by a conventional likelihood routine are still valid. On the other side the classic likelihood ratio tests can be performed in the same manner with the quasi-likelihood as with a likelihood function.

Klein and Spady (1993) show that the maximum quasi-likelihood estimator is asymptotically efficient under the assumption of independence of the random component of the latent variable, u_i, and the regressors.

2.3.3 Semiparametric Least Squares Estimator

Ichimura (1993) introduces a semiparametric estimator of the SIM (2.1) which is based on a least squares argument. More precisely he proposes to choose for $\hat{\beta}$ the value that maximizes in β the function

$$S_n(\beta) = \sum_{i=1}^{n} \{y_i - \widehat{F}_{hi}(x_i^T \beta)\}^2 \tag{2.10}$$

with $\widehat{F}_{hi}(x_i^T \beta)$ the estimate of the conditional expectation $E(Y_i|X_i^T \beta)$ given in (2.8). Here, also as in the case before, for theoretical reasons the target function (2.10) should include a trimming function to eliminate observations for which the kernel regression is imprecise. In practice the trimming can be ignored.

Maximization of (2.10) is performed iteratively using the usual methods of maximizing a nonlinear function.

Ichimura (1993) proves that under some regularity conditions the estimator resulting from maximizing (2.1) is asymptotically efficient and asymptotically normally distributed as Klein and Spady's (1993) estimator presented before. Consequently its covariance matrix is given by $(1/n)\Sigma$ with Σ as in (2.9).

The semiparametric least squares estimator and the quasi-likelihood estimator are asymptotically equivalent given that they converge to the same limiting law. Whether to use one or the other depends on which one of the target functions, (2.10) or the quasi-likelihood, is easier to maximize.

2.4 The Example Revisited

Table 2.1 shows the normalized estimates for the coefficients β in the four models introduced in chapter 1. The first column contains the parametric logit fit for each of the four data sets already shown in chapter 1. The other columns show the semiparametric fit using respectively the Klein and Spady (1993) method and weighted average derivative estimation. As before the quasi-likelihood estimates were determined using Gauss with a maximization routine which was gently provided by Bo Honore and Ekaterini Kyriazidou and the WADE estimates were obtained with XploRe 3 for bandwidth $h = 0.8$. All the coefficient estimates are normalized from being divided by the absolute value of the coefficient estimate of the first regressor.

The estimates are not much different except for the logit in Klein and Spady

	M.L.	Klein-Spady	WADE
		Logit	
intercept	0.7870		
	(0.1093)		
X1	-1	-1	-1
	(0.12235)	(0.0976)	(0.1145)
X2	1.4601	1.3967	1.4802
	(0.1413)	(0.1055)	(0.1089)
		Logit with a bump	
intercept	0.4933		
	(0.1557)		
X1	-1	-1	-1
	(0.1669)	(0.1389)	(0.1795)
X2	1.3317	1.6267	1.2326
	(0.1733)	(0.2570)	(0.1754)
		Logit with heterosc.	
intercept	0.5162		
	(0.1704)		
X1	-1	-1	-1
	(0.1809)	(0.1571)	(0.1882)
X2	1.1725	1.0304	1.1453
	(0.1821)	(0.1518)	(0.1846)
		CLL	
intercept	1.2772		
	(0.1321)		
X1	-1	-1	-1
	(0.1220)	(0.1990)	(0.1074)
X2	1.5647	1.6436	1.5477
	(0.1571)	(0.2896)	(0.1156)

Table 2.1: *Results of the Parametric and Semiparametric fits. Standard errors inside brackets.*

with bump where the semiparametric estimate for X_2 is a little closer to the true value 2. This result confirms the intuition stated before with the analysis of the plots in Figure 1.8 and is coherent with the result of Ruud (1983) to be discussed below. The estimates for the standard deviations are also all rather near and show that all coefficients are non null.

Ruud (1983) studies the properties of the maximum likelihood estimator for discrete responses when the link function is misspecified. The author concludes that in this situation the maximum likelihood procedure can provide consistent estimates for the slopes coefficients up to a scaling factor if

$$E(X_i|X_i^T\beta) = x_i^T\beta \quad i = 1, \cdots, n$$

When the regressors are jointly normal distributed as in the experiments treated here, the above condition is verified. Ruud (1986) extends this result for a more general behavior of the regressors. That is, by determining appropriate transformations of the sample data points in order that the regressors verify the above condition the author defines a weighted M-estimator which gives consistent estimates up to a factor of proportionality for the slope coefficients of the index function even when the link function is misspecified.

2.5 Confidence Bands

In last section it was pointed out that in certain conditions the parametric and semiparametric estimates of the index function in a binary choice model do not differ much. However, in those situations the parametric estimate of the link may deviate from the link estimate in the semiparametric model. In the semiparametric model the link is estimated using only very few assumptions given the index. To check the parametric assumptions one may assess how significantly the parametric fit deviates from the semiparametric fit. If the parametric model is correctly specified both fits should be close. Statistically one can consider both fits are close if the parametric fit lies inside of a confidence band of the semiparametric fit and consequently conclude that the parametric model is well specified.

This section is devoted to the determination of uniform confidence bands for the semiparametric fit. It shows how they can be used as a tool to check the specification of a binary choice model. Figure 2.6 is an example of this procedure showing the parametric fit (thick line) and the semiparametric uniform confidence band. The data were generated from a logit with a bump and were estimated parametrically assuming a wrong specification, the plain logit. Clearly the confidence bands together with the semiparametric estimate of the link (line with circles) reveal the existence of the bump.

The literature on confidence limits for the regression function is not very abundant and is mainly centered on the nonparametric approach. Härdle (1990) gives an exhaustive description about the construction of pointwise confidence intervals and uniform bands on nonparametric regression showing some applications. Still in the context of nonparametric regression Härdle and Marron (1991) use the wild bootstrap to calculate simultaneous confidence bars, Härdle and Bowman (1988) use bootstrap to deduce pointwise confidence limits and Rodríguez-Campos and Cao-Abad (1983) introduce a bootstrap procedure specifically for discrete choice models in order to built pointwise confidence intervals. Horowitz (1993) deduces semiparametric uniform confidence bands to asses the specification of a random-coefficients probit model for binary responses.

Uniform confidence bands have the advantage of allowing a global evaluation of the curve given that the coverage probability is defined globally all over the region where the bands are calculated. Thus, uniform confidence limits are more conservative than pointwise limits. Pointwise limits are usually easier to calculate but they allow only a local evaluation of the curve. For this reason only uniform bands are focused in this work.

The advantages of using confidence bands to check the link specification are the easy–to–understand insight and the easy visual evaluation of the parametric model compared to the alternative in a way that permits to judge if the difference between both is statistically significant or not. In case of misspecification it can also provide an hint of the type of deviation from the null which may lead in a possible reformulation of the parametric model as is suggested by Figure 2.6.

The remainder of the chapter will describe how to construct semiparametric uniform confidence bands. To exemplify the different procedures applications will be made with two simulated data sets introduced before – the logit model and the logit with a bump. The true index function is the same in both cases. The first data set illustrates the situation where the parametric model is well specified while the second illustrates a case of misspecification given that the data is estimated parametrically assuming the classic logit model (thus without incorporating the bump).

The construction of semiparametric confidence bands requires as a first step an estimate of the coefficients β of the index function, $\hat{\beta}$ to aggregate the multidimensional X variable. Then, all the techniques will be pursued using the fitted index.

To estimate β in the first step a $\sqrt{n}$-consistent estimator should be used. Horowitz (1993) uses the estimator under the null, that is the maximum likelihood estimator for the parametric link, which has the advantage of being easily calculated. However if the model is misspecified this estimate can be inconsistent. In this case it is not clear that the kernel estimate of the link based on a inconsistent projection of the index is going to be accurate or at least to capture the deviation of the true link from the assumed parametric link.

Another option is to estimate β with a semiparametric estimator which is still $\sqrt{n}$-consistent under the semiparametric alternative model for example the estimator of Klein and Spady (1993) or Delecroix and Hristache (1994). As stated before semiparametric estimators cannot identify the scale and the location (intercept) of the index $x^T\beta$. Consequently, a normalization of the vector of parameters is usually enforced. Scale and location will be incorporated in the estimate of the link.

A practical problem arises when comparing graphically the parametric with the semiparametric fit on the index projected under the semiparametric alternative. The plot draws the conditional probabilities against the fitted index. For each observation the parametric and the semiparametric index will be different (even when the same normalization of the coefficients is imposed in both). This difference may create a distortion on the plot.

One way of avoiding the kind of distortions mentioned in the paragraph above is by means of the procedure described in the algorithm 2.1.

1. Estimate β with a semiparametric estimator obtaining $\hat{\beta}$. Calculate the fitted index $v_i = x_i^T\hat{\beta}$, $i = 1, \cdots, n$.
2. Estimate the scale and the location of the parametric model given the fitted index v_i,$i = 1, \cdots, n$. This corresponds to estimate by maximum likelihood the parameters c and s on the model
$$P(Y_i = 1|V_i = v_i) = F(c + sv_i),\ i = 1, \cdots, n.$$
3. Plot the parametric fit $F(\hat{c} + \hat{s}v_i)$ together with the semiparametric fit and respective confidence limits against v_i, $i = 1, \cdots, n$.

Algorithm 2.1: *How to plot the parametric fit against the semiparametric fitted index*

Despite the more complexity in calculations it seems reasonable to use the index estimated semiparametrically to safeguard accuracy in case that the assumed parametric model is ill specified. However, in the examples presented

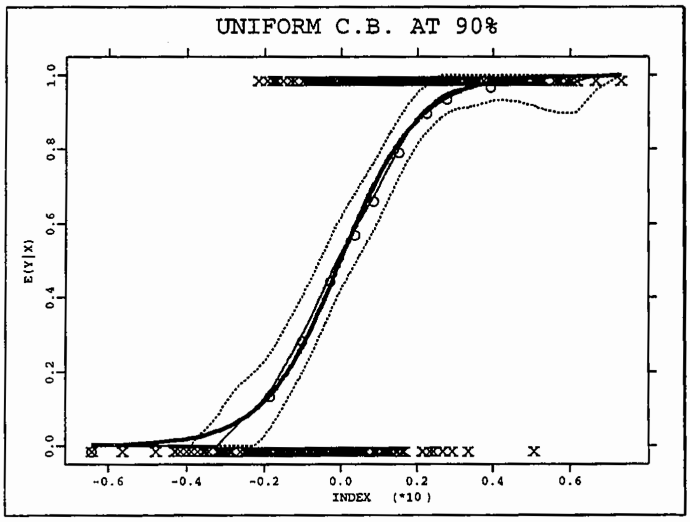

***Figure 2.3:** Uniform confidence bands. Calculations were done with the parametric fitted index. The data was generated from a logit link.*

here there is not a significant difference between both strategies which can be explained by the results of Ruud (1983, 1986) mentioned before.

There are basically two ways of constructing uniform confidence bands. One, consists in calculating pointwise confidence bars in a very fine grid of the projected index and correct the level with the Bonferroni correction in order to have simultaneous coverage. The drawback of this approach is that often the intervals may be quite wide specially if the grid has many points. The reason is that the correction to obtain global coverage probability assumes independence of the curve estimates between each point. Thus, to have global coverage of $1-2\alpha$ over a grid of np points the Bonferroni method corrects the level for each point in order to be equal to $1-2(\alpha/np)$.

Another method considers $\widehat{F}_h(v)-F(v)$ a stochastic process and bases a uniform confidence band on the asymptotic Gaussian distribution of $\sup_v |\widehat{F}_h(v)-F(v)|$. This is the approach taken by Horowitz (1993).

Horowitz (1993, p. 60) derives this distribution for $\widehat{F}_h(v)$ computed semiparametrically according to (2.2) with $\hat{\beta}$ the parametric estimate. However the Nadaraya-Watson estimator has an asymptotic bias proportional to the second moment of the kernel, the derivative of the regression function and the marginal density of x. In practice the bias is hard to estimate. Härdle (1990) proposes

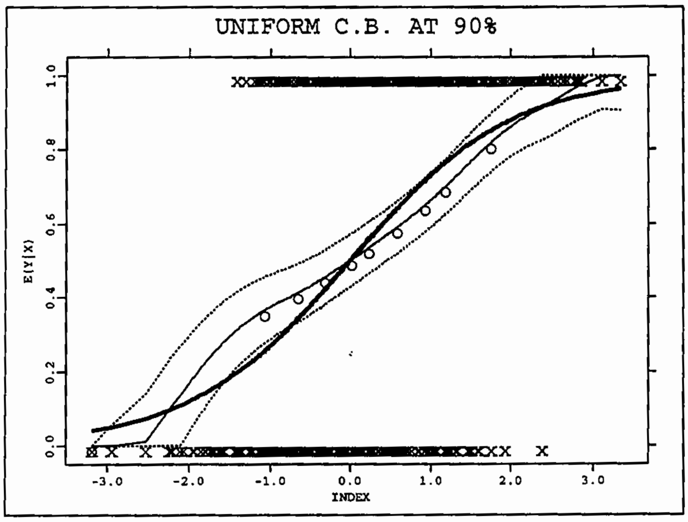

Figure 2.4: *Uniform confidence bands. Calculations were done with the parametric fitted index. The data was generated from a logit link with a bump.*

two approaches to deal with the bias. One is to choose h, the bandwidth, proportional to $n^{-1/5}$ times a sequence that tends slowly to zero in order that the bias vanishes asymptotically. The other uses bootstrap to estimate the bias.

A more direct and easy method to correct the bias in the Nadaraya–Watson estimator is the one proposed by Bierens (1987). Applying Bierens' result to the SIM gives (Bierens, 1987, p. 110)

$$\sqrt{nh}\{\widetilde{F}(x^T\beta) - F(x^T\beta)\} \xrightarrow{\mathcal{L}} N\left(0, \frac{\sigma^2(x^T\beta)}{p(x^T\beta)} C_K\right) \tag{2.11}$$

with $\sigma^2(x^T\beta) = V(Y|X^T\beta)$, $p(x^T\beta)$ the density of $X^T\beta$ at the point $x^T\beta$, C_K as

$$C_K = \int_{-\infty}^{\infty} K(u)^2 du. \tag{2.12}$$

and

$$\widetilde{F}(x^T\beta) = \left\{\widehat{F}_h(x^T\beta) - \left(\frac{h}{s}\right)^2 \widehat{F}_s(x^T\beta)\right\} \Big/ \left\{1 - \left(\frac{h}{s}\right)^2\right\}. \tag{2.13}$$

Here $\widehat{F}_t(x^T\beta)$, $t = h, s$ is the classic Nadaraya–Watson estimator in (2.2) for bandwidth t. Moreover $h = cn^{-1/5}$, $s = cn^{-\delta/5}$ with $c > 0$ and $0 < \delta < 1$.

The uniform confidence band incorporating Bierens' bias correction is calculated based on the following conjecture which combines Horowitz's and Bierens' results.

$$\lim_{n\longrightarrow\infty} P\left\{\sqrt{0.4\delta\log n}\left[\sqrt{nh}\sup_v\left\{\frac{\sigma^2(v)}{\hat{p}(v)}\right\}^{-1/2}|\widetilde{F}(v)-F(v)|\right.\right.$$
$$\left.\left.-d_n\right]<z\right\}=\exp\{-2\exp(-z)\} \quad (2.14)$$

where $1<\delta<5/3$ and

$$\begin{aligned} d_n &= (0.4\delta\log n)^{1/2}+(0.4\delta\log n)^{-1/2}\log\{C_K^*/(2\pi^2)\}^{1/2} \\ C_K^* &= \frac{2}{C_K}\int_{-\infty}^{\infty}K'(u)^2du \end{aligned} \quad (2.15)$$

while $\hat{p}_h(v)$ is the kernel density estimate of $p(v)$ given by

$$\hat{p}_h(v)=(nh)^{-1}\sum_i^n K\left(\frac{v-X_i^T\hat{\beta}_n}{h}\right), \quad (2.16)$$

and the other variables have the same meaning as before.

The confidence band is given by

$$\widetilde{F}(v_i)\pm\left\{\frac{c_\alpha}{(0.4\delta\log n)^{1/2}}+d_n\right\}\frac{C_K^{1/2}\hat{\sigma}(v_i)}{\{nh\hat{p}_h(v_i)\}^{1/2}}, \qquad i=1,\cdots,n \quad (2.17)$$

with $v_i=x_i^T\hat{\beta}_n$ the fitted index for the i-th observation and

$$\hat{\sigma}(v_i)=\widehat{F}_h(v_i)\{1-\widehat{F}_h(v_i)\}, \quad (2.18)$$

Algorithm 2.2 calculates the uniform confidence bands with Bierens' correction according to the procedure described above. This algorithm is implemented in XploRe 3 on the procedure `UNIFBAND` which code is given in the appendix.

Uniform confidence bands incorporating Bierens' bias correction are shown in Figures 2.3 and 2.4 for the index estimated parametrically and Figures 2.5 and 2.6 for the index estimated with the estimator of Klein and Spady (1993).

As before the particular estimate of the index does not produce significant differences in the shape of the confidence limits.

The semiparametric estimate (line with circles) deviates clearly from the parametric estimate (thick line) and induces the shape of the bump of the true link that generated the data.

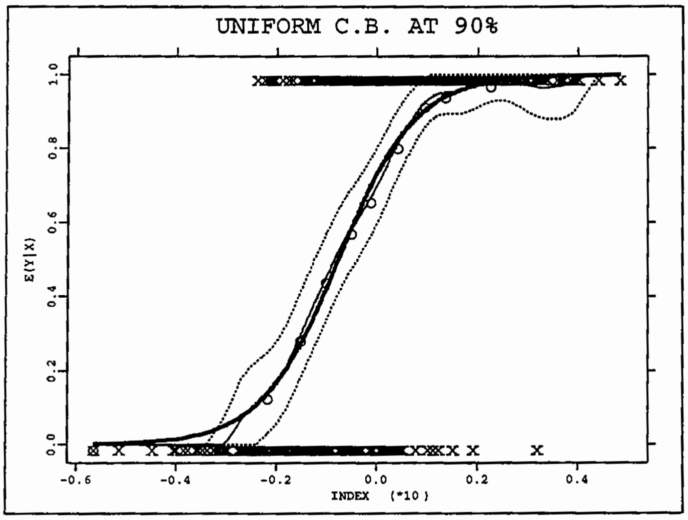

***Figure 2.5:** Uniform confidence bands. Calculations were done with the semiparametric fitted index. The data was generated from a logit link.*

Before finishing this section a remark should be done. The construction of the semiparametric confidence limits is bandwidth sensitive. Oversmoothing should be specially avoided because it can lead very easily to false rejections. Undersmoothing shows too much structure, e.g. little bumps, in the link. Heuristically the link is expected not to be very wiggly. Automatic procedures of bandwidth selection like least–squares cross validation have a tendency to undersmooth the data. A heuristic procedure of bandwidth choice can be taken instead. It consists on the following. Try different bandwidths and choose the smallest that doesn't give a rough wiggly curve. This was the conduct taken in the examples shown in this section. In much empirical applications with simulated data this procedure proved to give satisfactory results.

2.6 The Travel Mode-Choice and Credit-Scoring Applications

Let us return again to the applications introduced on the mode choice for travelling and credit-scoring.

Table 2.2 presents the semiparametric estimates of the coefficients of the in-

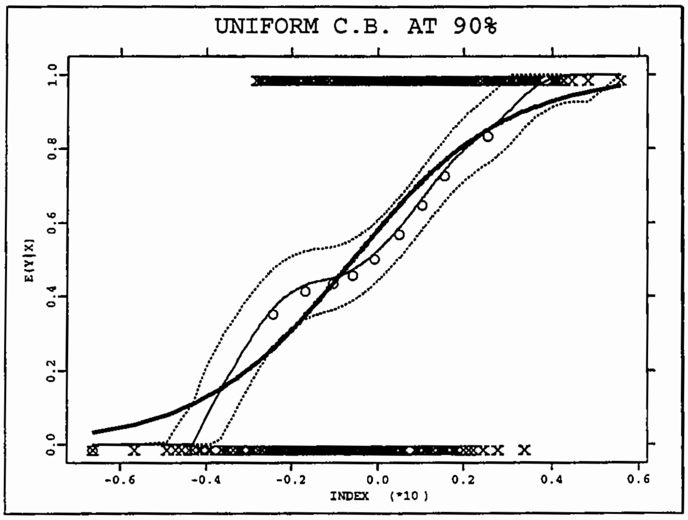

Figure 2.6: *Uniform confidence bands. Calculations were done with the semiparametric fitted index. The data was generated from a logit link with a bump.*

dex function and the respective estimated standard deviations for the problem of mode-choice for travelling in the way to work obtained with the maximum quasi-likelihood estimator. These estimates were determined using Gauss. As before the probability of travel mode-choice has as explanatory variables: number of cars owned by the travelers household, transit out-of-vehicle travel time minus automobile out-of-vehicle travel time (identified in the table by out-of-vehicle travel time), transit in-vehicle travel time minus automobile in-vehicle travel time (identified in the table by in-vehicle travel time) and transit fare minus automobile travel cost (identified in the table by travel cost). The regressors have been standardized to have mean zero and standard deviation equal to one. This transformation is convenient for the calculation of the semiparametric estimates in the maximization of the quasi-likelihood. The same table includes also parametric estimates assuming a logit model for the standardized regressors verifying the same normalization as the semiparametric estimates in order to be comparable. The parametric estimates were obtained with XploRe 3.

The semiparametric estimates are close to those obtained from the logit model. Figure 2.7 shows the parametric fit together with the semiparametric fit. The logit fit is plotted against the normalized parametric fitted index to be in the same scale as the semiparametric fit. The link in the semiparametric fit was estimated by kernel regression according to (2.2) with bandwidth $h = 0.8$.

1. Estimate the coefficients β in the index function, $\hat{\beta}_n$.
   Calculate the fitted index $v_i = x_i^T \hat{\beta}_n$, $i = 1, \cdots, n$.

2. For all v_i obtained in step 1, $i = 1, \cdots, n$:
   2.1 calculate $\widehat{F}_h(v_i)$ by kernel regression on $x_j^T \hat{\beta}_n$ with bandwidth h according to (2.2)
   2.2 Calculate $\widehat{F}_s(v_i)$ in the same way using bandwidth s.
   2.3 With $\widehat{F}_h(v_i)$ and $\widehat{F}_s(v_i)$ construct the bias corrected $\widetilde{F}(v_i)$. Use the Bierens' correction defined in (2.13).

4. Choose α in order to have a confidence level of $1 - \alpha$.
   Set c_α such that $\exp\{-2\exp(-c_\alpha)\} = 1 - \alpha$.

5. For all v_i obtained in step 1, $i = 1, \cdots, n$:
   5.1 Calculate the confidence limits
   $$\widetilde{F}(v_i) \pm [\{c_\alpha/(0.4\delta \log n)^{1/2}\} + d_n] \times \{C_K^{1/2}\hat{\sigma}(v_i)\}/\{nh\hat{p}_h(v_i)\}^{1/2},$$
   with C_K, $\hat{p}_h(v_i)$ and $\hat{\sigma}(v_i)$ given respectively by (2.12), (2.16) and (2.18), d_n as defined in (2.15) and $1 < \delta < 5/3$.

6. Plot the parametric regression and the confidence bands against the fitted index $x_i^T \hat{\beta}_n$ $i = 1, \cdots, n$.

Algorithm 2.2: *Semiparametric uniform confidence bands with Bierens' correction*

The semiparametric fit deviates from the logit essentially in the tails of the conditional probability curve.

Table 2.3 shows the semiparametric estimates for the coefficients of the index function with respective estimated standard deviations for the credit-scoring application obtained also with maximization of the quasi-likelihood. As before the probability of being a risky client has as explanatory variables: duration of credit in months, payment of previous credits (0 for good and 1 for bad), amount of credit, percentage of monthly payment in the individual income and age of the client. The regressors have been standardized for the reasons pointed out before. The same table includes also parametric estimates assuming a logit model with the standardized regressors for sake of comparison. Here too, parametric estimates and semiparametric estimates are not much

	semiparametric fit		logit fit	
	coeffic.	st. errors	coeffic.	st. errors
Intercept			1.4848	0.1023
Number of cars	1.0000	0.1174	1.0000	0.0972
Out-of-vehicle travel time	0.1672	0.0586	0.3114	0.0933
In-vehicle travel time	0.0685	0.0518	0.0826	0.0841
Travel cost	0.2671	0.0448	0.3199	0.0717

Table 2.2: Results of the semiparametric fit and logit fit on the problem of mode choice for travel.

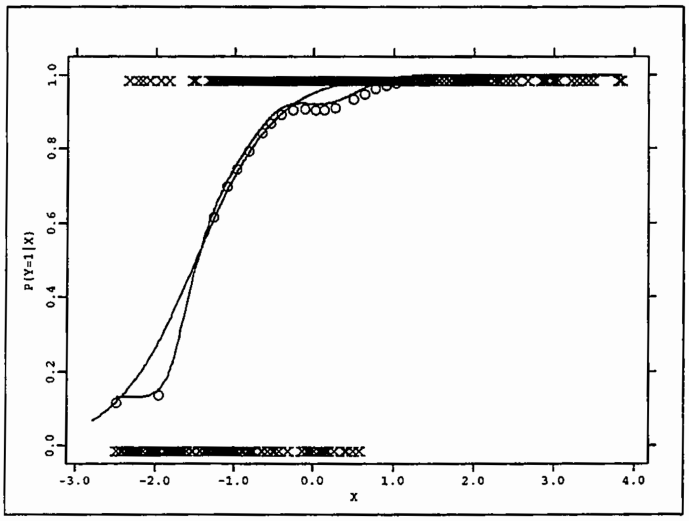

Figure 2.7: Semiparametric fit and Logit fit of the probability of individual choice of car for travel to work.

different.

Figure 2.8 shows the semiparametric fit together with the logit fit. Again the logit fit was plotted against the parametric fitted index with the same normalization as the semiparametric fitted index to be both in the same scale. As before the semiparametric fit was obtained by kernel regression with bandwidth $h = 3$. The semiparametric fit diverges considerably from the logit fit in the upper tail due mainly to the lack of observations in that region. However in the region where observations are concentrated the semiparametric fitted link de-

	semiparametric fit		logit fit	
	coeffic.	st. errors	coeffic.	st. errors
Intercept			-3.0144	0.2434
Duration of credit	1.0000	0.1593	1.0000	0.3093
Payment of previous credits	1.0247	0.1468	1.2047	0.2201
Amount of credit	0.8296	0.1148	0.5808	0.3231
Monthly payment	0.2305	0.0611	0.7629	0.2686
Age	-0.6405	0.1006	-0.7983	0.2581

Table 2.3: Results of the semiparametric fit and logit on the problem of credit-scoring.

viates slightly from the logit showing the existence of bumps in the probability curve.

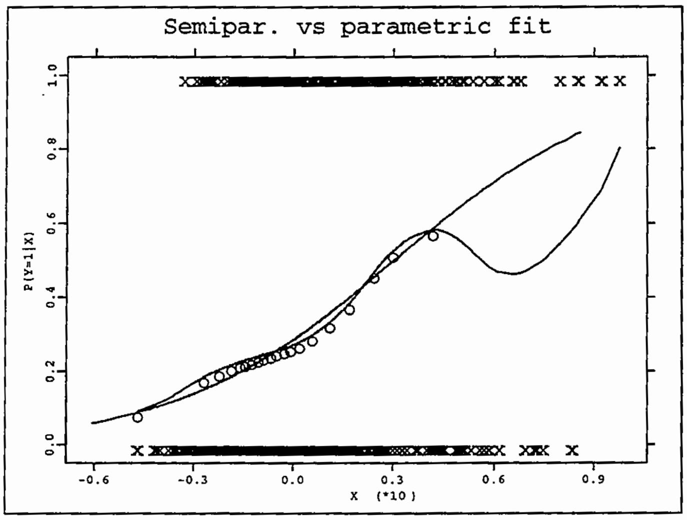

Figure 2.8: Semiparametric fit and Logit fit of the probability of a client to be a potential risk.

Figures 2.9 and 2.10 show an uniform confidence band for the semiparametric fit for respectively the travel mode-choice probability curve and the credit scoring. The bandwidths used to calculate the kernel regression were respectively $h = 0.8$ and $h = 3$. A logit fit on the normalized semiparametric fitted index is also plotted to assess the specification of the logit link according to

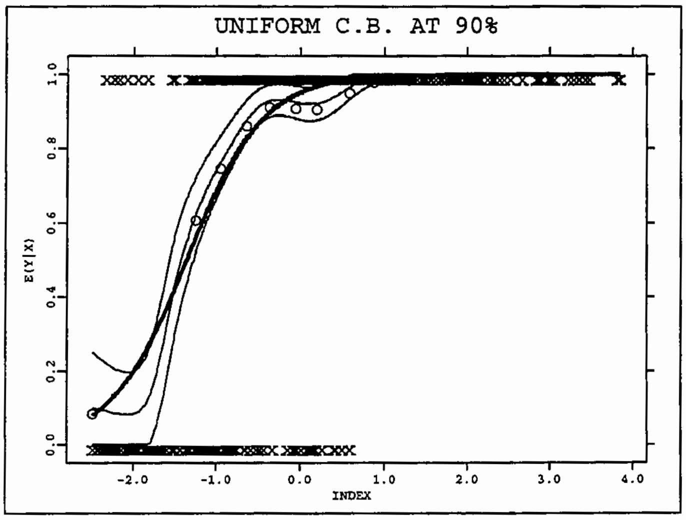

Figure 2.9: *Uniform confidence bands for the probability of individual choice of car to travel to work. Calculations were done with the semiparametric fitted index.*

the procedure described in section 2.5.

In both examples is not clear that the logit fit lies completely inside the confidence bands raising some suspicions about the correctness of the logit specification in these problems. This issue will be addressed in further chapters.

2.7 Concluding Remarks

An important question faced by the econometrician in modeling binary dependent variables is which specification to choose. One point is clear, between logit and probit the choice is irrelevant.

The choice of a specific model depends on the information and knowledge of the econometrician about the problem and data in question. However, it would be convenient to have a way of checking the adequacy of the chosen specification with the data to safeguard from ill-specified models that can produce inconsistent estimates and seriously erroneous predictions.

Heterogeneity on the behavior of the decision makers is likely to occur especially in cross-section models. As it was pointed out before, some of the more used parametric models neglect the effect of heterogeneity like the classic

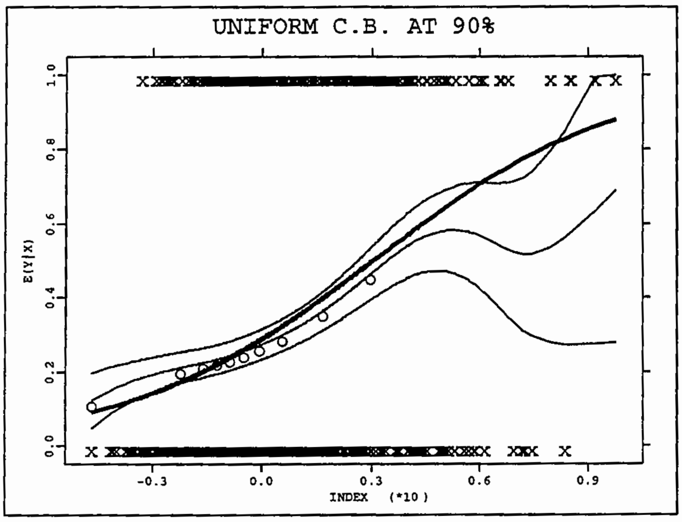

Figure 2.10: *Uniform confidence bands for the probability of a client to be a potential risk. Calculations were done with the semiparametric fitted index.*

probit or logit. More generally, it may be difficult to characterize in practice the distribution of the non-observed behavior of individuals. Therefore, usually it is assumed that it has a normal or a logistic form and a misspecification error may be present. This thesis deals with testing procedures that aim to detect the kind of misspecification just mentioned in a parametric model.

One can check a parametric model by comparing it with other parametric alternative. For example, one can implement a log likelihood ratio test to compare a probit model with a conditional probit model because they are nested. Indirectly the conditional probit and a logit are also compared (given the similarities between probit and logit). However this kind of test detects misspecification in a very restricted set and one has to have a very clear idea about the structure of the possible deviation from the null, that is the alternative model.

Chesher (1984) introduces a specification test to detect neglected heterogeneity in a parametric model. Heterogeneity is conceived as leading to variation in parameters assumed constant to all decision makers. Thus, heterogeneity results in models with random coefficients.

The author derives a score test for the hypothesis that the variances of the parameters of interest are zero in the case where the parametric model (without heterogeneity) is well specified. The test statistic is derived without specifying a particular distribution for the parameters of interest which has to verify only a general set of regularity restrictions (see Chesher, 1984). Therefore the set of models in the alternative hypothesis is a wide set in which the RCP is a particular element. The author shows that the test statistic turns out to be the information matrix test introduced by White (1982).

This work follows a different approach. The idea is to have a testing procedure to detect any sort of deviation of the link function leading to a wide class of alternative models. When there is very little information about the kind of possible deviations from the assumed parametric model this can be a useful strategy. The kind of test just mentioned can detect arbitrary heterogeneity when it brings out a misspecification of the assumed parametric link function.

The remain of this thesis studies and develops some recent existent tools allowing to test the specification of the stochastic structure of a parametric binary choice model in a wide set of deviations from the null. This corresponds to detect unspecified deviations of the link function keeping the index $x_i^T\beta$ fixed in order to avoid the curse of dimensionality and is equivalent to test the specification of a parametric model against a single index model. As it was said before, this alternative allows for heterogeneity of the form of heteroscedasticity as a function of the index.

Most of the econometricians and statisticians favor parametric models mostly because they are easy and ample interpretable in corresponding to economic theory. Often they are also easily to estimate than semiparametric or nonparametric models. However, a parametric model that is not correctly specified according to the information transmitted by the data may be misleading the true nature of the population behavior and may lead to inconsistent inferences.

Returning to the examples introduced in section 1.8 and more precisely to the plot in the lower left of Figure 1.8, observe that there is a systematic deviation from the true curve that has generated the data (the logit with a bump) and the misspecified parametric fit assuming a logit. This example shows it is important to have tools that assess a possible misspecification of parametric models. These tools can also be extended as means to test the adequacy of economic theory (see as example Hildenbrand, Härdle and Jerison, 1991). On the other hand, they may also provide a guidance to the definition of a new specification for the model or to give a guidance for the reformulation of the theory behind.

This thesis proceeds with the analysis of methods to assess the specification of a binary choice model. It will focus predominantly on semiparametric techniques. One possible approach evaluates the specification of a parametric model based on the construction of confidence bands for a semiparametric fit and was already studied in this chapter. The other is by means of hypotheses tests and will be developed in the remainder of this thesis.

Chapter 3

The HH–test

3.1 Introduction

Chapter 1 introduces some parametric binary choice models like the logit, the probit, the complementary log-log and the conditional probit. The binary choice model can be represented as $P(Y = 1|X = x) = F(x^T\beta)$ and was characterized to have two distinct components: the link function $F(\bullet)$ and the index function, $X^T\beta$. The index was interpreted as the deterministic part of the latent variable in (1.5) or the deterministic part of individual utilities which reflect the average behavior over all individuals. The link was understood to be determined by the distribution function characterizing the latent variable or the distributional properties of stochastic utilities. Note also that the link can be viewed simply as a nonlinear regression function operating on the index, a linear function that aggregates a set of explanatory variables.

Misspecification of the parametric binary choice model can be due to a wrong choice of the link or an ill-specified index function. This work is concerned with misspecification on the link function or on parametric distributional properties assumed in the model. From here on it will be assumed that the index has a linear specification according to what was asserted in the antecedent chapters. Note that the linearity of the index is a very popular assumption. However, the procedures in study allow for parametric specifications other than linear.

Assuming correct a linear form for the index function then misspecification can still occur due to omitted relevant variables, like interaction terms, or wrong transformations of the included variables. Collet (1991) develops several graphical tools to deal with these problems. Another approach is contemplated on the framework of Generalized additive models (GAM) of Hastie and Tibshirani (1990).

GAM models consist on a parametric fixed link operating on an additive function which is a sum of a finite number of univariate functions, that is, each acting in one explanatory variable. These functions are unknown and are estimated nonparametrically. Thus, GAM allows for nonparametric transformations of the explanatory variables of the index function. Turlach (1994) develops testing techniques to allow variable selection in GAM. If one defines exhaustively the set of explanatory variables and then apply these techniques can have still a parsimonious model and a low risk of omitting relevant variables.

As principle followed in this thesis the index is assumed to be well specified. Therefore, the analysis of the tools mentioned above is not in the scope of this work. In practice the following procedure can be pursued. Before testing the specification of the link function perform an analysis of the index function for the assumed link using the tools mentioned in the precedent paragraphs in order to choose a correct specification of the index. It is important to remark however that it is not yet well known how the functional form of the link influences the index specification and vice versa. Therefore, if the link is reformulated one should verify if the index is still well specified given the new specification of the link function. Experience says that usually it is. If not, one has to reformulate the index. Then, the link should be tested again and the all procedure should be repeated till the link or the index specification are accepted.

Chapter 2 has introduced a more flexible formulation for the binary choice model, the SIM. Flexibility arrives from substituting the fixed parametric link by an unknown function keeping the parametric index. The resulting model is semiparametric.

The semiparametric model safeguards from misspecification on the link. However, parametric models have their own recognized virtues. Namely, estimation and inference are easier and analysts like to know the form of the functions they deal with for interpretation purposes or to derive related models (for example by calculating analytical derivatives). Therefore, the point of view of this work is to retain the parametric specification of the binary choice model but having means of checking its adequacy (which may eventually guide on a possible reformulation) to shelter from ill-specified models.

A first procedure to assess the correctness of the link was introduced in chapter 2 with the construction of a semiparametric uniform confidence band.

Here, the aim is to evaluate the specification of the link function by means of a formal hypotheses test.

3.1.1 The Problem

Suppose a set of observations $\{y_i, x_i\}$ which are realizations of the random variables (Y_i, X_i), $i = 1, \cdots, n$. Assume that these observations are independent and identically distributed to the random variable (Y, X) where X is a k-dimensional vector assuming values in $I\!R^k$ and Y is a binary variable assuming only the values 0 or 1. Presume that the data were estimated assuming the parametric model

$$P(Y = 1|X = x) = E(Y|X = x) = F(x^T\beta)$$

with $F(\bullet)$ a parametric link function which may assume one of the particular forms presented in chapter 1.

The aim is to know whether the data are adequately fitted by the parametric model $F(x^T\beta)$ given the index function. This amounts to test for any deviations of the functional form of the link $F(\bullet)$ assuming a given (correct) specification of the index. In this case a semiparametric test should be performed according to

$$\begin{aligned} H_0 &: E(Y|X = x) = F(x^T\beta) \\ H_1 &: E(Y|X = x) = H(x^T\beta) \end{aligned} \tag{3.1}$$

with $H(\bullet)$ an unknown continuous function that assumes values between 0 and 1. Note that the model under the alternative hypothesis is the semiparametric model studied in chapter 2. Note also that the semiparametric test is a test of nested hypotheses in the sense that the alternative includes the null hypothesis.

The semiparametric test searches for misspecification in a wider set than a parametric test, that is a test where the alternative model is a parametric one. The later reduce to the comparison between two parametric models. On the other hand, semiparametric tests detect misspecification in a more restrict set than nonparametric tests where the alternative is a nonparametric model. The later detect any kind of deviations of a parametric model and can be viewed as goodness-of-fit tests.

To perform the semiparametric test in the context of binary choice model according to (3.1) the test procedure of Horowitz and Härdle (1994) also known as the HH–test can be used. This test procedure is very appealing given its easy implementation and fruitful motivation. The HH–test can be applied to a variety of models to which the binary choice model is only a particular case. This chapter introduces the HH–test on the context of binary choice models. First, a brief review of some existing testing techniques that can be applied to the problem in study and are somehow related with the HH–test is presented below.

3.1.2 Specification Tests for Binary Choice Models

The HH–test is inspired in the framework of Conditional Moment (CM) tests introduced in the seminal work of Newey (1985). CM specification tests constitute an important framework that inspired most of the actual test procedures regardless of being parametric, semiparametric or nonparametric.

CM tests are based on the estimation of a set of moment conditions that have to be verified when the null hypothesis is true and violated under the alternative. The motivation of CM tests comes from the property verified in models with exogenous variables in which the expectation of the score vector of an observation, conditioned on the exogenous variables, is zero. This implies that functions of exogenous variables should be uncorrelated with the elements of the score vector, and suggests that a specification test could be performed based on sample covariances of functions of exogenous variables and the score function evaluated at the maximum likelihood estimator. Newey (1985) constructed such a statistic and found a Chi-square asymptotic distribution under the null with degrees of freedom equal to the number of moment conditions evaluated. The well known White (1982) specification test based on the information matrix equivalence theorem can be considered as an example of a CM test.

By choosing adequately the precise moment conditions to evaluate one obtains statistics for consistent testing against specific alternatives like heteroscedasticity, omitted variable or non-normality. Newey (1985) gives interesting examples of this procedures for the probit model.

Semiparametric specification tests in binary response models constitute a recent field of research and only a few approaches are known. The Bayesian proposals of Czado (1992) and Chappell, Czado and Newton (1993) are examples. However, the Bayesian approach of qualitative choice models is not in the scope of this work.

Another proposal can be found in Whang and Andrews (1993). The authors present a test statistic with a Chi-square asymptotic distribuiton which is very much based on the idea of Newey (1985) C.M. tests but it is generalized in order to allow for an infinite-dimensional estimator like a nonparametric density function or regression function estimator.

The authors use their test procedure to test distributional assumptions like normality which can be applied in particular to binary choice models. This is equivalent to test a probit model under the null against a SIM according to the semiparametric test (3.1). The test procedure is based on the evaluation of the first order conditions of a semiparametric estimator for the parameters of the index function, like the estimators of Ichimura (1993) or Klein and Spady (1993), at the parametric estimate obtained under the null, the probit model in this example. If the parametric model is correctly specified then those first order conditions evaluated at the probit estimate should be very close to zero and consequently the test statistic should be also close to zero.

The authors claim that in presence of misspecification if the parametric estimator is not consistent the test is usually consistent. However, empirical experience shows that most of the time semiparametric estimates of the coefficients of the index do not differ significantly from parametric estimates. As example see the results obtained in table 2.1 in chapter 2. One reason pointed there for this behavior is developed by Ruud (1983). If parametric and semiparametric estimates of the index are close one can expect that the power of Whang and Andrews (1993)'s test will be very poor in finite samples. On the other hand usually the first order conditions of the semiparametric estimator are complicate to compute making difficult to calculate the test statistic.

Turlach (1994) introduces also a test procedure that allows to perform the semiparametric test (3.1) in order to assess the correctness of the link function. The test statistic evaluates the difference between the weighted average derivative estimate of the index and a convenient scaled parametric estimate obtained under the null. Again as in the precedent test procedure it is necessary that both estimates differ significantly for the test to have power in finite samples. For the reasons pointed before this will not be always the case. Turlach (1994) identifies some situations characterized for certain local alternatives against which the test has no power.

The test techniques of Whang and Andrews (1993) and Turlach (1994) have in common the fact that they appraise misspecification in the link function by a certain evaluation of the discrepancies between the parametric estimate of the coefficients of the index function and the semiparametric one. Those tests require that the semiparametric and the parametric estimates are considerably different in order to have power in finite samples. In practice this may not happen as it was pointed out before which makes the practical performance of those tests very doubtful and turns preferable the use of other testing strategies like the HH–test.

The motivation behind the HH–test seems to be more appealing. Deviations on the parametric link are assessed by evaluating the difference between the parametric estimate of $P(Y = 1|X^T\hat{\beta})$ and the respective nonparametric estimate (given the same estimate of the index, $X^T\hat{\beta}$, which has to be consistent under both hypotheses), weighted by the residuals of the parametric regression. This is very much the idea behind the construction of confidence bands pursued in chapter 2 to test the correctness of the parametric regression.

One can assess the specification of the parametric binary choice model by a nonparametric test procedure. Assuming that the index is correctly specified then if the test possibly detects deviations from the null they are due to an ill specification of the link function.

Recently some nonparametric testing techniques have been developed that somehow have inspired the HH–test, reason that makes them worth to mention here in the following.

Bierens (1990) is based on a CM condition that is verified when the parametric model is correct and violated when some unspecified deviation from the parametric model occurs. Proença (1993) shows a comparison between the power in finite samples of Bierens (1990)'s test and the HH–test with a simulation study for some binary choice models. The study reveals that the HH–test performs better to test a parametric model against the semiparametric alternative of (3.1).

Azzalini, Bowman and Härdle (1989) generalize the likelihood ratio test in order to compare a parametric model with a nonparametric alternative. The asymptotic distribution of the test statistic is very difficult to derive and the authors advise the use of bootstrap to deduce the critical values. Bootstrap makes the test computationally demanding.

Härdle and Mammen (1993) use a different approach. They evaluate the statistical significance of a L2–distance between the parametric model and a nonparametric estimate and deduce the asymptotic distribution of the test. However, the asymptotic bias and variance of the test need big sample sizes to be accurately estimated. Therefore, the authors advise the use of *wild bootstrap* instead. As before, the implementation of the test is computationally burden.

A goodness-of-fit test specifically constructed for binary regression models was introduced by le Cessie and van Houwelingen (1991). The test aims to detect all deviations from a hypothesized parametric model. It is based on a statistic entailing nonparametric smoothing of the standardized residuals of the parametric fit. Smoothing the residuals of the parametric model instead of the data has the advantage of avoiding the typical bias of nonparametric estimation because under the null the residuals have conditional expectation zero. Although the authors do not deduce explicitly the asymptotic distribution of the statistic, they claim based on the results of simulations that it can be well approximated by a given scaled Chi-square distribution. To improve the power of the test in finite samples they obtain a finite sample correction for the bias and variance of the statistic when the null is a logit model. The approach of le Cessie and van Houwelingen (1991) was important to inspire the improvements to the HH–test to be proposed in chapter 6.

Gozalo (1993) introduces a test statistic to compare a parametric or semiparametric model against a nonparametric alternative. The test statistic is based on the square of the difference between the model estimated under the null and its nonparametric estimate which is then divided by the variance of the latter. The resulting statistic has a Chi-square limiting distribution. Gozalo (1993)'s procedure requires samples with many observations not only to guarantee the accuracy of the nonparametric estimator of the regression function and the estimate of the respective variance but also because the calculation of the statistic needs the splitting of the sample.

Semiparametric tests are advantageous over nonparametric tests because they elude the so-called *curse of dimensionality* which means that the rate of convergence is getting slower when the number of explanatories increases and the typical lack of power on finite samples that may arrive to nonparametric test procedures. The idea behind the semiparametric approach is to restrict the set of alternatives in order to gain power in finite samples but still test against a large set of misspecification. Dimensionality is reduced by aggregating the multidimensional explanatory variable with a parametric index function.

3.2 The HH–test Statistic

In this section the HH–test statistic is introduced in two steps. First, it will be assumed that the index function is totally known, that is, the coefficients vector β is known and is not necessary to be estimated. This approach is unrealistic but theoretically has some important features that have to be distinguished. In the second step β is unknown and has to be estimated. This is the usual situation in practice. It will be seen that estimation of β brings some distortions in finite samples relatively to the ideal asymptotic behavior of the statistic affecting its performance.

3.2.1 Known Index Function

Suppose that the index function is completely known, that is, the coefficients β are known. Define $v_i = X_i^T\beta$ the known index function and $r(v_i)$ the residuals of the parametric regression obtained as $r(v_i) = Y_i - F(v_i)$.

The HH–test evaluates the difference between the model with the parametric assumed link $F(v_i)$ and a respective nonparametric estimate, call it in the moment $\widehat{F}(v_i)$. This estimate can be based on the Nadaraya-Watson kernel regression defined in (2.2). The test statistic is obtained based in the difference between both functions weighted by the residuals of the parametric regression, that is will be function of $r(v_i)\{\widehat{F}(v_i) - F(v_i)\}$. When the parametric model is misspecified $F(v_i)$ and $\widehat{F}(v_i)$ should differ in a large amount not explained only to random sampling error making the statistic significantly different from zero. On the other hand, if the parametric model is correctly specified both functions should be near and the statistic should assume values close to zero.

One condition for the above reasoning to be valid is that the nonparametric regression $\widehat{F}(v_i)$ is asymptotically unbiased. Asymptotic unbiasedness of kernel regression estimators can be achieved with Bierens' bias correction defined in (2.13) within chapter 2. On the other hand, for technical reasons (which will be analyzed later in the continuation of the section) $\widehat{F}(v_i)$ has to be independent of Y_i. This goal can be attained with the LOO estimator introduced in chapter 2 in equation (2.8). In the sequel the nonparametric regression of the dependent variable on the index function with LOO estimation and Bierens' bias correction will be denoted by $\widetilde{F}_i(\bullet)$.

Finally the test statistic is given by,

$$T_n = \sqrt{h}\sum_{i=1}^{n} u(v_i)\{Y_i - F(v_i)\}\{\widetilde{F}_i(v_i) - F(v_i)\} \qquad (3.2)$$

where h is the bandwidth used in the kernel regression and $\widetilde{F}_i(v_i)$ the LOO estimate with Bierens' bias correction, obtained according to,

$$\begin{aligned}\widetilde{F}_i(v) &= \left\{\widehat{F}_{hi}(v) - \left(\frac{h}{s}\right)^2 \widehat{F}_{si}(v)\right\} \Big/ \left\{1 - \left(\frac{h}{s}\right)^2\right\} \qquad (3.3)\\ \widehat{F}_{ti}(v) &= \frac{\sum\limits_{j\neq i}^{n} K\{(v-v_j)/t\}\, y_j}{\sum\limits_{j\neq i}^{n} K\{(v-v_j)/t\}}, \qquad t = h, s\end{aligned}$$

Here $h = cn^{-1/5}$, $s = cn^{-\delta/5}$ with $c > 0$, $0 < \delta < 1$. Observe that $\widetilde{F}_i(v)$ is a linear combination of two LOO estimates, $\widehat{F}_{hi}(v)$ and $\widehat{F}_{si}(v)$, calculated for different bandwidths respectively h and s. On the other hand, $u(v_i)$ is a non-negative weight function that down-weights extreme observations.

Suppose that $\widehat{F}_h(v_i)$, the simple Nadaraya-Watson in (2.2) is used on the calculation of the HH-statistic instead of $\widetilde{F}_i(v_i)$. Then, if the bandwidth tends to zero, the sum $\sum_{i=1}^{n} u(v_i)\{Y_i - F(v_i)\}\{\widehat{F}_h(v_i) - F(v_i)\}$ approaches $\sum_{i=1}^{n} u(v_i)\{Y_i - F(v_i)\}\{Y_i - F(v_i)\}$, which has a finite, positive expectation. In practice, this implies that the mean of this version of the statistic depends strongly on the bandwidth, which is undesirable. To arrange that T_n has zero mean regardless of the bandwidth it is necessary to assure that the two factors $\widehat{F}_h(v_i) - F(v_i)$ and $Y_i - F(v_i)$ are independent for all i. This is accomplished when LOO estimate is used on nonparametric regression and will be called as independence correction. This reasoning is supported by the results of the simulations performed and to be presented later in next chapter.

Horowitz and Härdle (1994) have shown that under H_0 and under some suitable regularity conditions a central limit theorem of Hall (1984) holds and T_n is asymptotically distributed as a $N(0, \sigma_T^2)$ with,

$$\begin{aligned} \sigma_T^2 &= 2C_K \int_{-\infty}^{\infty} u(v)^2\{\sigma^2(v)\}^2 dv \\ C_K &= \int_{-\infty}^{\infty} K(u)^2 du \\ \sigma^2(v) &= V[Y|X^T\beta = v] \end{aligned} \tag{3.4}$$

3.2.2 The HH–test as a CM test

The HH–test can be formulated as a CM test according to the framework of Newey (1985).

When the parametric model is correctly specified then

$$E\{Y - F(X^T\beta)|X = x\} = 0$$

Thus, functions of explanatory variables should be uncorrelated with the residuals $Y - F(X^T\beta)$. This has suggested Newey (1985) to deduce a specification test statistic based on the evaluation of the moment conditions

$$E\{M_j(Y, X, \beta)\} = E[\{Y - F(X^T\beta)\}w_j(X, \beta)] = 0 \qquad j = 1, ..., q \tag{3.5}$$

where $w_j(X,\beta)$ $j = 1, ..., q$ are weighting functions chosen in a way such that at least one of the q conditions in (3.5) fails when the alternative hypothesis is true. When the model is correct clearly by the law of iterated expectations the condition (3.5) is true. Because the test is based on the verification of conditional moment restrictions it is called a conditional moment (CM) test. By choosing adequately the weight functions $w_j(\bullet)$ one obtains statistics for consistent testing against specific alternatives. See Newey (1985) for examples of parametric alternatives and Bierens (1990) for a nonparametric alternative test.

The HH–test can be formulated as a CM test because the test statistic (3.2) may be derived based on a moment condition similar to conditions in (3.5) for a particular choice of the weight function $w(X,\beta)$. Given that the aim of the test is to assess the discrepancies between the parametric model and the alternative semiparametric SIM, or equivalently detect any deviation from the parametric link, it is natural to suggest for $w(X,\beta)$ the following expression

$$w(X,\beta) = u(X^T\beta)\{H(X^T\beta) - F(X^T\beta)\},$$

where $u(X^T\beta)$ is a non-negative weight function and $H(\bullet)$ is unknown. The moment condition becomes

$$E[\{Y - F(X^T\beta)\}u(X^T\beta)\{H(X^T\beta) - F(X^T\beta)\}] = 0 \qquad (3.6)$$

If the parametric model is correctly specified the above moment condition is clearly true. Under the alternative H_1, that is $E(Y|X) = H(x^T\beta)$ the HH–test will be powerful if the condition (3.6) is violated. Therefore, the HH–test will work when the parametric link function is misspecified if the following condition

$$E[\{H(X^T\beta) - F(X^T\beta)\}u(X^T\beta)\{H(X^T\beta) - F(X^T\beta)\}] \neq 0,$$

is verified. Thus, to keep the power of the test the weight $u(v)$ has to be chosen such that $E[u(X^T\beta)\{H(X^T\beta) - F(X^T\beta)\}^2]$ be positive.

The HH-statistic (3.2) estimates the moment condition (3.6) by a sample moment using the available observations $\{y_i, x_i\}$, $i = 1, \cdots, n$. However, this condition includes the function $H(\bullet)$ which is unknown. When calculating the test statistic it has to be substituted by a consistent estimate, a nonparametric kernel regression of $E(Y|X^T\beta)$ on $X^T\beta$ denoted by $\widetilde{F}_i(\bullet)$ which, for the reasons already explained, includes also independence correction by LOO estimation and bias correction by Bierens' technique.

3.2.3 The HH–statistic Under the Alternative

Horowitz and Härdle (1994) derive the distribution of the HH–test statistic under local alternative hypotheses. They consider the sequence of alternatives H_n defined by,

$$H_n[x^T\beta] = F[x^T\beta] + n^{-1/2}h^{-1/4}\Delta_n[x^T\beta],$$

where Δ_n, $n = 1, 2, ...$ is a sequence of uniformly bounded functions that converges uniformly to a limit function $\Delta(v)$.

Under the sequence of alternative models H_n the HH–test statistic (3.2) is asymptotically distributed as a $N(\mu, \sigma_T^2)$ with $\mu = E\{u(X^T\beta)\Delta^*(X^T\beta)^2\}$ with $\Delta^*(X^T\beta) = \Delta(X^T\beta) - F'(X^T\beta)X^T\gamma$ where $\gamma = \lim n^{-1/2}h^{-1/4}(\hat{\beta} - \beta)$ when n goes to infinity. Hence the HH–test statistic has a positive asymptotic mean under alternatives whose distance from the null is $O(n^{-1/2}h^{-1/4})$ and therefore has power on this set of alternatives.

The HH–test and the nonparametric test of Härdle and Mammen (1993) are equivalent in one-dimensional problems in the sense that they detect misspecification for models with the same rate of distance to the parametric model. The benefits of the semiparametric test are visible for higher-dimensional problems because it will keep the same one-dimensional rate of distance while the nonparametric test suffers from the *curse of dimensionality*, that is the rate becomes slower with the increase of dimension.

To conclude, under many alternatives, there will be systematic differences between $\widetilde{F}_i(v_i)$ and the assumed $F(v_i)$. Given that Horowitz and Härdle (1994) proved that T_n has positive mean when $E[Y_i|v_i]$ is not $F(v_i)$ then T_n can be used to construct a one-sided test for the adequacy of a specified link function which rejects for large values. The one-sided test improves the power of the test in finite samples.[1] For large enough sample sizes this test rejects if $T_n/\sigma_T > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1-\alpha$ percentile of a standard normal distribution.

3.2.4 Estimated Index Function

In practice the linear index function has to be estimated and β has to be replaced in the calculation of the statistic by a consistent estimate $\hat{\beta}$. Consider $\hat{v}_i = X_i^T\hat{\beta}$. In this situation the statistic is given by

$$T_n = \sqrt{h}\sum_{i=1}^{n} u(\hat{v}_i)\{Y_i - F(\hat{v}_i)\}\{\widetilde{F}_i(\hat{v}_i) - F(\hat{v}_i)\} \qquad (3.7)$$

[1] I am grateful to Joel Horowitz for this remark

The independence correction works as long as the index v_i is known. If the coefficients in the index have to be estimated, $\widetilde{F}_i(\hat{v}_i) - F(\hat{v}_i)$ and $Y_i - F(\hat{v}_i)$ become dependent, since $F(\hat{v}_i)$ and $\widetilde{F}_i(\hat{v}_i)$ depend on Y_i via the estimation of the model parameters. This produces a non-zero finite sample expectation of the test statistic under the null hypothesis; as the simulations in this thesis suggest, this bias remains important even for samples of thousand of observations.

One condition for the test to work asymptotically is that the estimator of the index coefficients is $\sqrt{n}$-consistent. Therefore it converges faster than the nonparametric estimate of the link function and consequently asymptotically the statistic behaves like in the situation of known linear index.

Call $T_n(v)$ to the HH-statistic calculated for the known index according to (3.2) and $T_n(\hat{v})$ to the HH-statistic calculated at the estimated index as in (3.7). Horowitz and Härdle (1994) try to prove that asymptotically $T_n(\hat{v})$ has the same behavior as $T_n(v)$ and thus is distributed under the null according to a $N(0, \sigma_T^2)$ with σ_T^2 as in (3.4). Unfortunately their proof is not totally correct[2] which makes the above result not valid.

In their proof Horowitz and Härdle (1994) consider an arbitrary nonstochastic sequence $\{\beta_n\}$ in $\mathbb{R}^k$ verifying $\sqrt{n}(\beta_n - \beta) = O(1)$. Define $\tilde{v} = X^T\beta_n$ and call $T_n(\tilde{v})$ to the HH-statistic calculated at $X_i^T\beta_n$. The authors claim that $T_n(\hat{v}) - T_n(v) = o_p(1)$ if $T_n(\tilde{v}) - T_n(v) = o_p(1)$ because β_n is arbitrary. However, this is not completely true but $T_n(\hat{v}) - T_n(v) = o_p(1)$ if $sup_{\beta_n} T_n(\tilde{v}) - T_n(v) = o_p(1)$.

To construct the test statistic it is necessary to define $u(v)$ and to obtain an estimate of $\sigma^2(v)$. Technical reasons force $u(v)$ to be continuous, independent of the sample and with support contained within that of $X^T\beta$. In practical applications the last restriction is the most important. One suggestion is to consider $u(v)$ identically equal to one in its support and zero otherwise. The authors remark that depending on how σ_T^2 is estimated, using a support of $u(v)$ that exceeds the support of $X^T\beta$ may cause a substantial positive bias for this estimate and a corresponding loss of power. They suggest to define the support of $u(v)$ based on the fitted values $x_i^T\hat{\beta}$ excluding the 5% or 2.5% biggest and the 5% or 2.5% smallest values. This conduct proved to be reasonable in the simulations of Proença (1993) with binary choice models.

For estimating $\sigma^2(v)$ a consistent estimator under H_0 has to be used that should not become excessively large under H_1 in order to avoid a loss of power. For binary response models $\sigma^2(v)$ is a well known function of $F(v)$ and can be

[2] I am grateful to Alois Kneip who detected the error and gave also some hints on how to fix it.

	HH-stat	p-value		HH-stat	p-value
h = 0.2	1.49	0.93	h = 1.6	1.15	0.88
h = 0.4	2.30	0.99	h = 1.8	0.58	0.72
h = 0.6	3.09	1.00	h = 2.0	0.04	0.52
h = 0.8	3.42	1.00	h = 2.2	-0.47	0.32
h = 1.0	3.09	1.00	h = 2.4	-0.59	0.16
h = 1.2	2.45	0.99	h = 2.6	-1.53	0.06
h = 1.4	1.78	0.96	h = 2.8	-2.10	0.02

Table 3.1: ***Results of the HH–test on the mode-choice to travel data set. The HH–statistic divided by its estimated standard deviation and respective p–value.***

estimated under the alternative by

$$\hat{\sigma}^2(v) = \widetilde{F}_i(X_i^T\hat{\beta})\{1 - \widetilde{F}_i(X_i^T\hat{\beta})\}.$$

Algorithm 3.1 calculates the HH–test. This algorithm is implemented in XploRe 3 on the procedure **HHTEST** which is included in the Appendix.

The semiparametric hypothesis test on performing the analysis conditional on the index $X^T\beta$ detects misspecification for models $E(Y|X^T\beta = v) = F(v)$. Nevertheless it can detect also misspecification for the broad model $E(Y|X) = F(x^T\beta)$ under certain conditions. To be more specific suppose that the later model is not correct and the true specification of $E(Y|X = x)$ is $m(x)$. Horowitz and Härdle (1994) show that the HH–test is consistent if $E_{X|v}m(X) \neq F(v)$ where $E_{X|v}$ denotes the expectation relative to the distribution of X conditional on $X^T\beta = v$, on a subset of the support of $w\{x, \beta\}$ that has positive probability. Therefore, the test is inconsistent to detect any kind of misspecification if $P\{m(x) = F(x^T\beta)\} < 1$ but the expectation $E_{X|v}\{m(X)\} = F(v)$ is verified almost everywhere in the support of $u(\bullet)$.

3.3 The Travel Mode-Choice and Credit-Scoring Applications

In this section the HH–test is applied to the data sets analyzed in last chapters, that is the travelling mode choice on the way to work and credit-scoring, to test the adequacy of the logit fit. The HH–test is calculated according to

1. Estimate the coefficients β in the index function, by maximum likelihood with the parametric model obtaining $\hat{\beta}$.
   Calculate the fitted index $\hat{v}_i = x_i^T \hat{\beta}$, $i = 1, \cdots, n$.

2. For all $\hat{v}_i$ obtained in step 1, $i = 1, \cdots, n$:
   2.1 calculate $\widehat{F}_{hi}(\hat{v}_i)$ the LOO kernel regression with bandwidth h smoothing relatively to $\hat{v}_j = x_j^T \hat{\beta}$ according to (2.8)
   2.2 Calculate $\widehat{F}_{si}(\hat{v}_i)$ in the same way using bandwidth s.
   2.3 With $\widehat{F}_{hi}(\hat{v}_i)$ and $\widehat{F}_{si}(\hat{v}_i)$ construct the bias corrected $\widetilde{F}_i(\hat{v}_i)$. Use the Bierens' correction defined in (3.4).

3. Define the weight function $u(\bullet)$ to be equal to one for the 90% or 95% central values of the ordered $\hat{v}_i$, $i = 1, \cdots, n$.

4. Calculate the HH–statistic
   $$T_n = \sqrt{h} \sum_{i=1}^{n} u(\hat{v}_i)\{Y_i - F(\hat{v}_i)\}\{\widetilde{F}_i(\hat{v}_i) - F(\hat{v}_i)\}.$$

5. Calculate the estimate of the standard deviation of the statistic, $\hat{\sigma}_T$ by
   $$(2C_K/n) \sum_{i=1}^{n} \{u(\hat{v}_i)\}^2 [\widehat{F}_{hi}(\hat{v}_i)\{1 - \widehat{F}_{hi}(\hat{v}_i)\}]^2 / \hat{p}_{hi}(\hat{v}_i), \tag{3.8}$$
   with $\hat{p}_{hi}(\hat{v}_i)$ and C_K given respectively by
   $$\hat{p}_{hi}(v) = (nh)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\{(v - \hat{v}_j)/h\} \tag{3.9}$$
   and (2.12).

6. Choose the size of the test to be α. Set $c_{\alpha/2}$ and c_α equal respectively to the quantiles $1 - \alpha/2$ and $1 - \alpha$ of a standard normal.
   Accept the parametric model if
   - $|T_n/\hat{\sigma}_T| < c_{\alpha/2}$ for a two sided-test
   - $T_n/\hat{\sigma}_T < c_\alpha$ for a one-sided test

   Otherwise, accept that the parametric model is misspecified.

Algorithm 3.1: The HH–test

	HH-stat	p-value
h = 0.2	0.65	0.74
h = 0.4	0.78	0.78
h = 0.6	0.46	0.68
h = 0.8	0.05	0.52
h = 1.0	-0.76	0.22
h = 1.2	-1.78	0.03
h = 1.4	-2.78	0.00

Table 3.2: Results of the HH–test on the credit-scoring data set. The HH–statistic divided by its estimated standard deviation and respective p–value.

algorithm (3.1) using XploRe 3. Several values for the bandwidth h are tried. The results concerning the mode choice to travel data are in Table 3.1 while Table 3.2 contains the results for the credit-scoring data set.

The value of the HH–test statistic divided by its estimated standard deviation depends clearly on the bandwidth used. For both data sets it gives different conclusions on acceptance or rejection of the logit model depending on the bandwidth used. In general, when the bandwidth is small the HH–statistic tends to assume positive values while it tends to be negative for superior values of the bandwidth, decreasing with the increase of h.

Figure 3.1 and Figure 3.2 show the parametric logit fit and the kernel regression on the parametric index respectively for the mode-choice to travel and credit-scoring data. The kernel regression was calculated for several bandwidths to illustrate undersmoothed and oversmoothed fits.

For the credit-scoring the region where the parametric fit and kernel regression are more apart corresponds the the region where there are few observations which can affect the accuracy of the kernel regression there. Thus, it is not surprising that the HH–test doesn't give a rejection of the logit fit for bandwidths smaller than 1.2, though for $h = 0.4$ the p–value is high. For $h = 1.2$ and $h = 1.4$ the HH–test rejects the parametric model which can be due merely to oversmoothing the data as it is discussed below. In the travel mode-choice data the kernel regression deviates clearly from the logit fit. Thus, it is understandable that the HH–test leads to rejection for almost all bandwidths tried. These results are consistent with the conclusions reached in chapter 2 when using uniform confidence bands to check the specification of the logit fit.

In Figure 3.1 and Figure 3.2 it can be seen that undersmoothing makes

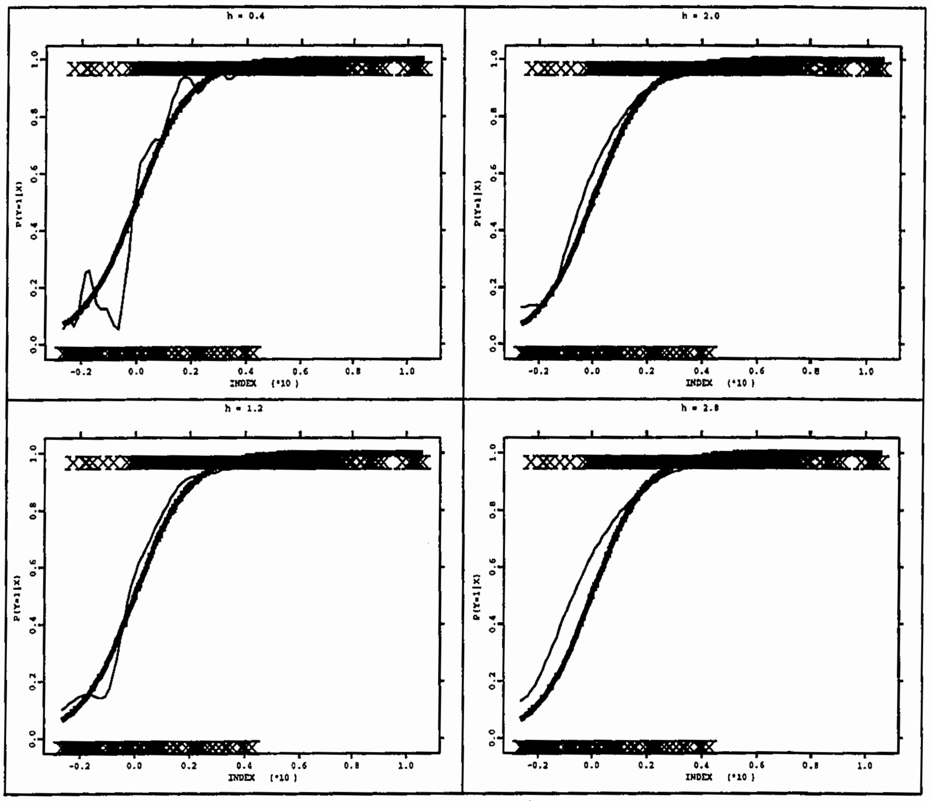

Figure 3.1: Mode choice to travel data: parametric regression and kernel regression. Upper left – h=0.4, lower left – h=1.2, upper right – h=2.0 and lower right – h=2.8

the kernel regression too wiggly. The HH-statistic tends to be high in this case because of the erratic behavior of the kernel regression. On the other hand, oversmoothing provokes a systematic behavior of the kernel regression which tends to be over the parametric fit in the region where observations for the response are mainly equal to zero (and consequently parametric residuals tend to be negative) and to be under the parametric fit when the parametric residuals tend to be positive. Therefore for observations in the first region the statistic is made by the product of mainly negative residuals times the difference between the kernel regression and the parametric fit which is positive while for observations in the second region this difference is negative but the parametric residuals are mainly positive. That explains why for bigger bandwidths the HH-statistic tends to be negative. When the bandwidth gets greater the kernel

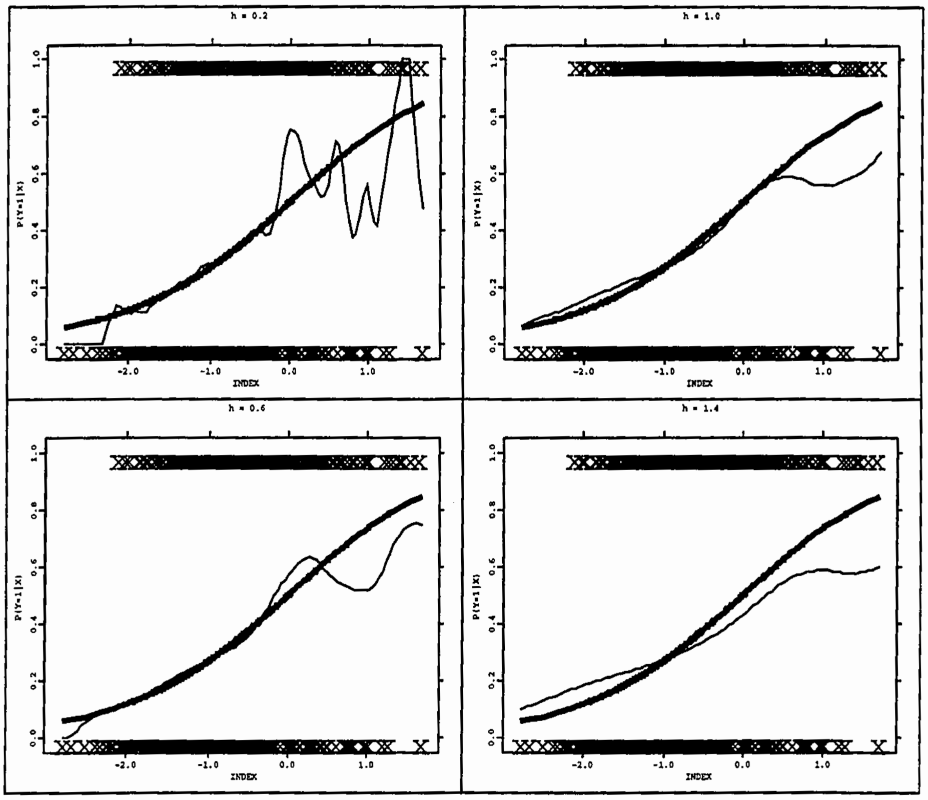

Figure 3.2:]

Credit-scoring data: parametric regression and kernel regression. Upper left – h=0.2, lower left – h=0.6, upper right – h=1.0 and lower right – h=1.4

regression deviates successively more from the parametric fit and therefore the HH–statistic tends to assume smaller negative values which can lead easily to rejection. Therefore, these examples suggest that oversmoothing can provoke severe undesirable effects on the behavior of HH–test leading to false rejections of the parametric model. This issue deserves to be further analyzed and will be subject of chapter 4.

3.4 Concluding Remarks

The HH–test appears to be an appealing tool to test the link specification in qualitative choice models. It combines an easy implementation with a direct and intuitive interpretation. The semiparametric approach avoids the *curse of dimensionality* and reduces the complexity of the problem. Before considering the HH–test an useful tool for the practioner, its behavior in finite samples has to be investigated.

Horowitz and Härdle (1994) do not explore the empirical finite sample properties of the HH–test for discrete choice models. Careful studies on this subject are presented by Proença (1993) and Proença and Ritter (1993). Their main goal is to assess the empirical size and power of the test in finite samples. It is important to know the relation between power and the dimension of the sample to have a benchmark on the sample size needed to get reliable results on rejection or acceptance of the assumed parametric specification. It is also important to evaluate the influence of bandwidth choice (for a given sample dimension) on the power and size. The data analyzed in this chapter suggest that oversmoothing may induce a pernicious behavior of the HH–test, namely leading to false rejections of the parametric model. These questions will be the subject of chapter 4.

Chapter 4

Finite Sample Properties of the HH–test

4.1 Introduction

This chapter aims to study the finite sample properties of the HH–test statistic for testing the specification of the link function in models with binary responses. This study will be done by simulations.

One important issue is to evaluate the effect of bandwidth choice on the properties of the test. The ideal would be to have a test robust to the choice of h under the null although it is very likely that h will always influence the detection of a specific alternative. In order to assess the influence of the bandwidth either in rejecting a true hypothesis or in rejecting a false hypothesis, all the experiments were implemented for a set of different bandwidths.

One main purpose is to study the empirical size and empirical power of the test in finite samples and to analyze the effect of the sample size on the accuracy of the test (that is to accept H_0 when H_0 is true and to reject it otherwise). In the experiment H_0 is chosen to be a logit model. The data are generated from several link functions: logit, logit with a bump, logit with heteroscedasticity and complementary log–log. Those models were introduced already in chapter 1. For the logit, H_0 is true. The others are included in the set of alternatives of the semiparametric test (3.1). They were chosen in order to have different type of possible deviations from the logit that may occur frequently in practice as it was pointed out in chapter 1.

When calculating the HH–test statistic two corrections have to be used: the independence correction by using the leave–one–out kernel regression and the Bierens' to correct the bias. Specially the last one is computationally expensive (it amounts to calculate two kernel regressions instead of one). The importance of those corrections on the achievement of the normal asymptotic behavior of the statistic will be a subject of study in this chapter.

As it was pointed out in chapter 3, to derive the limiting distribution of the HH–test statistic it is necessary to do an approximation that consist in substituting the statistic evaluated at the the fitted index by the statistic evaluated at the true index. Although evaluating the statistic at the true index or at the fitted index may be asymptotically equivalent, in finite samples some discrepancies may arise. In this case the performance of the test would be negatively affected due to a bad approximation of the critical values of the test by the adequate percentiles of the standard normal. For this reason the behavior of the statistic evaluated at the true index and the statistic evaluated at the fitted index will be analyzed and compared. The aim is to assess how significantly the empirical size and power of the test is affected by estimating the coefficients in the index function.

In calculating the HH–test statistic choices have to be made about the weight function and the specific δ for Bierens' correction in (3.4). In a former study of Proença (1993) these aspects were analyzed by simulations. Because they have turned not to be of main interest, their analysis will not be pursued here but referred to. In the mentioned work, it was concluded that the particular value chosen for δ was not important for the performance of the test. It was also concluded that considering the weight function defined equal to one for the 90% central values of the fitted index and zero otherwise is not damaging the performance of the test. This conduct has the advantage of choosing the weight function "automatically".

The plan of the chapter is as follows. Section 2 describes the general characteristics of the simulation study. Section 3 analyzes the repercussion of the independence and Bierens' corrections on the behavior of the statistic. Section 4 evaluates the effect of estimating the index function. Section 5 studies the performance of the empirical power and size of the HH–test for finite samples of growing size. Section 6 gives the concluding remarks.

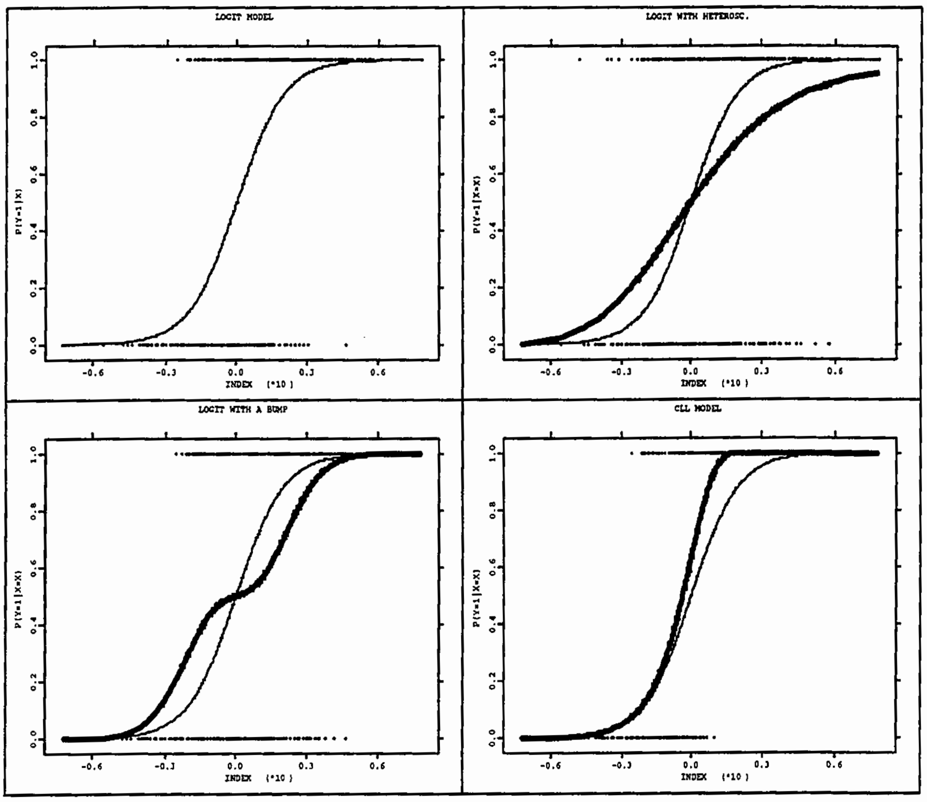

Figure 4.1: Upper left – the logit model, lower left – the logit and the logit with a bump, upper right – the logit and the logit with heteroscedasticity, lower right – the logit and the complementary log-log.

4.2 Description of the Experiment Design

This section describes the general characteristics of the simulation studies of this chapter.

The hypothesis H_0 considered in all studies is the logit model,

$$H_0 : E\{Y|X = x\} = \{1 + \exp(-x^T\beta)\}^{-1} = F(x^T\beta) \qquad (4.1)$$

Four different models are considered: the logit, the logit with a bump, the logit with heteroscedasticity and the Complementary-log-log. Their expressions are defined in respectively (1.23), (1.24) and (1.22). A plot with the shape of these models can be seen in Figure 4.1.

For each model 500 samples of n observations were generated. The data generation processes are described in chapter 1, section 1.8. All calculations were performed in GAUSS.

For the sample size n the values of 100, 200, 400 and 1000 are considered while the bandwidth h assumes the values $0.2, 0.4, 0.6, 0.8, 1, 1.2$.

For the calculation of the HH–test statistic the weight function $u(\bullet)$ is defined to be equal to one for all $x_i^T\hat{\beta}_n$ (or $x_i^T\beta$ when the statistic is evaluated at the true index) except for 5 percent at either extreme, where it is zero. The parameter δ used in Bierens' bias correction was fixed equal to 0.4. A previous study included in Proença (1993) shows that these choices are reasonable and not relevant for the performance of the test.

4.3 The Role of Independence and Bias corrections

This section studies the small and moderate sample effect of independence and Bierens' bias corrections on the achievement of the normal limiting distribution of the HH–statistic. The study also examines the effect of changing the bandwidth in the behavior of the statistic.

4.3.1 Description of the study

Four different versions of the statistic will be considered corresponding to the absence of both corrections, the absence of only bias correction, the absence of only independence correction and the HH–test statistic including both corrections. These versions of the statistic are given below.

$$T_n^h = h^{1/2}\sum_{i=1}^{n} u(x_i^T\beta)\{y_i - F(x_i^T\beta)\}\{\hat{F}_h(x_i^T\beta) - F(x_i^T\beta)\}, \qquad \text{(V1)}$$

with $\hat{F}_h(\bullet)$ as in equation (2.2),

$$T_n^h = h^{1/2}\sum_{i=1}^{n} u(x_i^T\beta)\{y_i - F(x_i^T\beta)\}\{\hat{F}_{ih}(x_i^T\beta) - F(x_i^T\beta)\}, \qquad \text{(V2)}$$

with $\hat{F}_{ih}(\bullet)$ as in (2.8),

$$T_n^h = h^{1/2} \sum_{i=1}^{n} u(x_i^T\beta)\{y_i - F(x_i^T\beta)\}\{\widetilde{F}(x_i^T\beta) - F(x_i^T\beta)\}, \qquad \text{(V3)}$$

where

$$\widetilde{F}(x_i^T\beta) = \left\{ \hat{F}_h(x_i^T\beta) - \left(\frac{h}{s}\right)^2 \hat{F}_s(x_i^T\beta) \right\} \Big/ \left\{ 1 - \left(\frac{h}{s}\right)^2 \right\},$$

with $s = hn^{(1-\delta/5)}$; and

$$T_n^h = h^{1/2} \sum_{i=1}^{n} u(x_i^T\beta)\{y_i - F(x_i^T\beta)\}\{\widetilde{F}_i(x_i^T\beta) - F(x_i^T\beta)\}, \qquad \text{(V4)}$$

with $\widetilde{F}_i(\bullet)$ as in (3.4) and s as before. These versions correspond to a classic Nadaraya–Watson estimator (V1), a leave-one-out Nadaraya–Watson estimator (V2), a Bierens' bias corrected Nadaraya–Watson estimator (V3), and a Bierens' corrected leave-one-out Nadaraya–Watson estimator (the full HH–statistic), respectively.

All this versions of the statistic were calculated for the true index function. This is the ideal situation, i.e., when the HH–test can attain the best performance, because no noise is introduced by the estimation of β. Thus a behavior close to the limiting normal distribution should be expected when the sample size grows. On the other hand, possible anomalies in this behavior due to the absence of the corrections in the statistic can be more purely identified.

The study proceeds as follows. Given a model for $F(x_i^T\beta)$ and a sample size n, 500 samples $(y_i, x_i)_g, i = 1, \cdots, n,\ g = 1, \cdots, 500$, are generated according to the description in section 4.2. Here only two models were considered for $F(\bullet)$, the logit model and the logit with a bump according respectively to (1.21) and (1.23). For each generated sample the versions of the statistic (V1), (V2), (V3) or (V4) are calculated for the different values of the bandwidth. The values of the statistic standardized by its asymptotic standard deviation estimated according to (3.8) are reported. The aim is to describe some characteristics of the statistic distribution for each version, bandwidth and sample size. Therefore the empirical mean, median, standard deviation, 5th percent and 95th percent percentiles were calculated based in the values for the standardized statistic (in each version) obtained in the 500 samples.

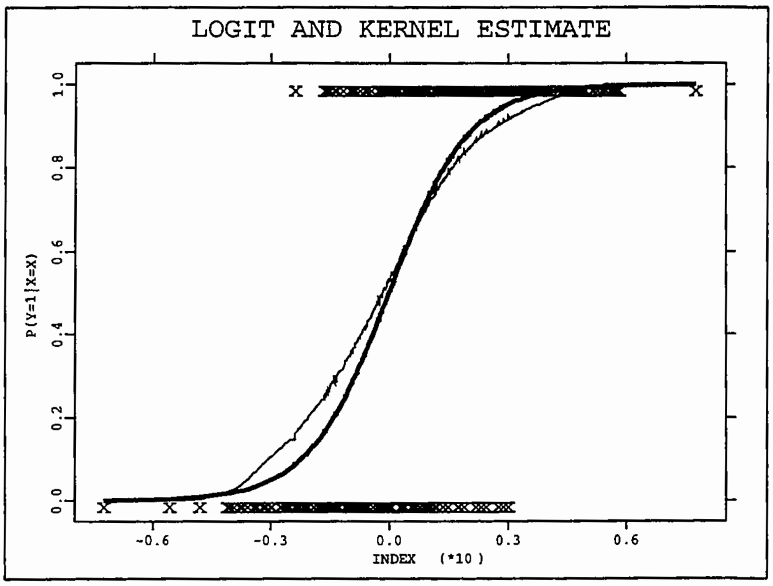

Figure 4.2: The logit link with the leave–one–out kernel estimate with $h = 1$.

4.3.2 Results

For the HH–test statistic with both corrections, given by (V4), the lower right plots in Figures 4.3, 4.4, 4.5 and 4.6 show the 5th, 95th percentiles and the mean of the standardized statistic calculated under H_0 for the bandwidths $h = 0.2, 0.4, \cdots, 1.2$ and the sample sizes 100, 200, 400 and 1000; together with 5th and 95th percentiles of the standard normal.

For 100 observations the percentiles of the statistic almost coincide with the respective percentiles of the standard normal and the mean of the statistic is almost zero for all the bandwidths considered. The variance is stable and close to one.

For 200, 400 and 1000 observations, one notices a fanning out of the variance which becomes more important as the size of the sample and the bandwidth grow. This feature becomes apparent for bandwidth larger than a certain level that depends on the sample size (e.g for $n = 200$ it happens for h larger than 0.8 while for $n = 1000$ it happens for h larger than 0.4). Possibly this is an effect of oversmoothing in the nonparametric estimation of the link also detected in chapter 3, section 3.3, when the HH–test was applied to the data sets of travelling mode-choice and credit scoring.

Figure 4.2 shows the logit link together with the kernel regression with bandwidth $h = 1$ for 400 data points. This bandwidth is oversmoothing the data with the result that the kernel estimate is flatter than the logit curve. This deviation due to oversmoothing is expressed by a high value of the statistic and the consequent divergent behavior of the percentiles. When the sample size is smaller oversmoothing will occur for bigger bandwidths in a way inversely related with the number of observations.

The results allow to conclude that the behavior of the HH–test statistic under the null with ideal conditions (evaluated at the true index) approaches very accurately the limiting normal critical values and mean even for small sample size given that the bandwidth is small enough.

The absence of Bierens' bias correction on the HH–test using variation (V3), produced the simulation results shown in the the lower left plots in Figures 4.3, 4.4, 4.5 and 4.6. The picture is very similar whether the bias correction is implemented or not. For the smaller sample size tried $n = 100$ there is a slight difference in the 95th percentile, that is, the absence of the bias correction slightly under-evaluates the respective percentile of the standard normal.

Omitting independence correction using versions (V2) or (V1) has a clear significant effect on the behavior of the test. The plots on the upper sides of Figures 4.3, 4.4, 4.5 and 4.6 show that without independence correction the statistic is very bandwidth dependent under the null and has a strong positive bias which increases with decreasing bandwidth.

The behavior of the four versions of the HH–test under the alternative logit with a bump is shown in Figures 4.7, 4.8, 4.9 and 4.10.

Omitting independence correction has the same effect as before as is shown in the upper plots of the referred Figures. The absence of Bierens' correction has a slight effect only for small sample sizes (100 and 200) resulting in a slight under evaluation of the 95th percentile.

For 100 observations the results are very disappointing. For all bandwidths the mean of the statistic is practically zero suggesting a total absence of power of the test. The mean of the HH–test statistic increases with the sample size as expected. It increases also with bandwidth suggesting that bigger bandwidths are more able to detect this kind of alternative. For $n = 400$ the mean of the statistic (on the right lower plot) overcomes the critical value of the test, the $95th$ percentile of the normal for h larger than 0.8 while for $n = 1000$ the mean is always bigger than the upper critical value for all bandwidths considered.

The study suggests that the independence correction is essential, while the

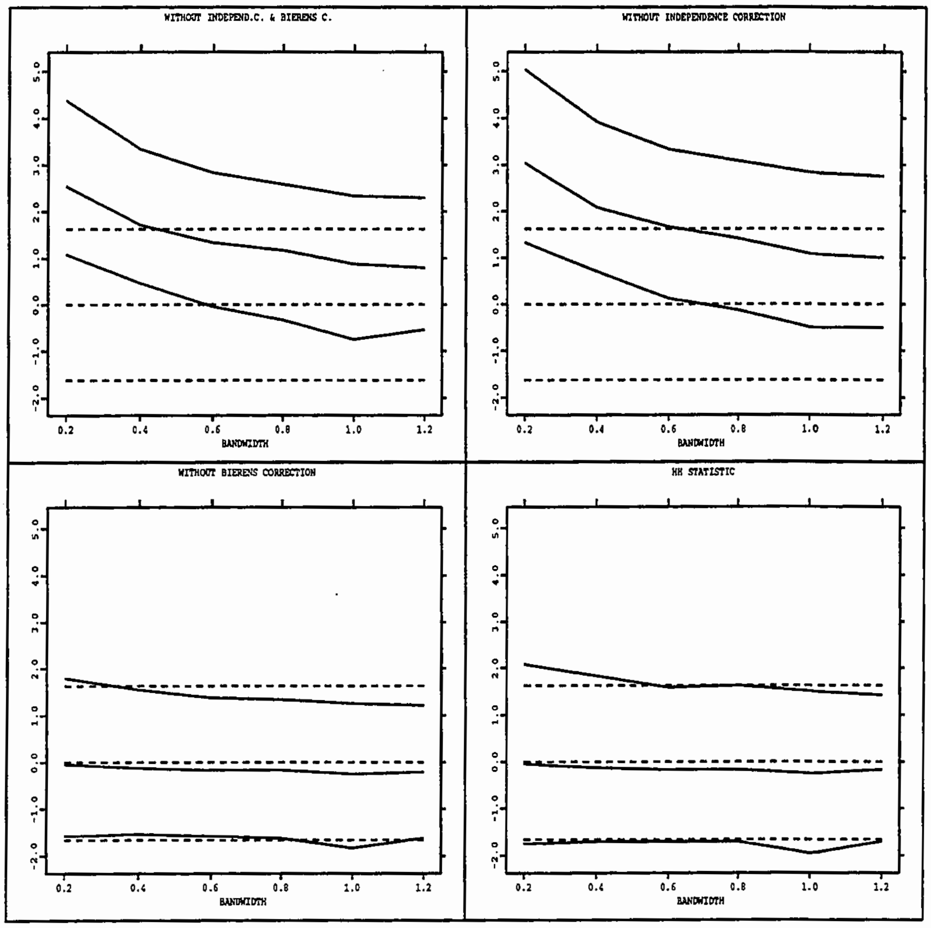

Figure 4.3: *Effect of bias and independence corrections.* 100 *observations, logit model, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

Bierens' correction, although required for theoretical reasons, has little practical importance for small to medium sample sizes. Under the null, convergence to the asymptotic distribution is rapid for small bandwidth. Large bandwidth cause problems. Under the alternative the test shows a notorious lack of power for small sample sizes. However for moderate sample sizes the results are already very satisfactory.

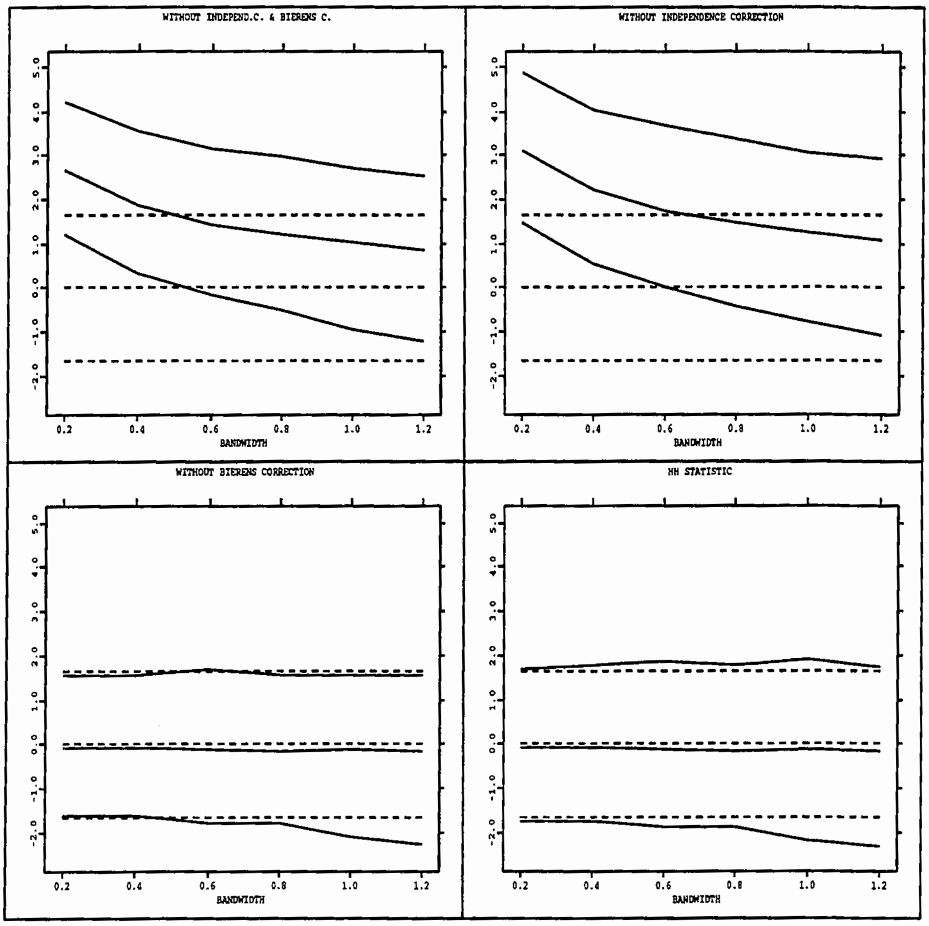

Figure 4.4: Effect of bias and independence corrections. 200 *observations, logit model true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

4.4 The Effect of Estimating the Index Function

The results in section 4.3 show that the HH–test statistic converges quickly to its limiting distribution under the null (provided that the bandwidth is not too large). Also, under the alternative analyzed the mean of the statistic overcomes the normal critical value of the test for moderate sample sizes. These results

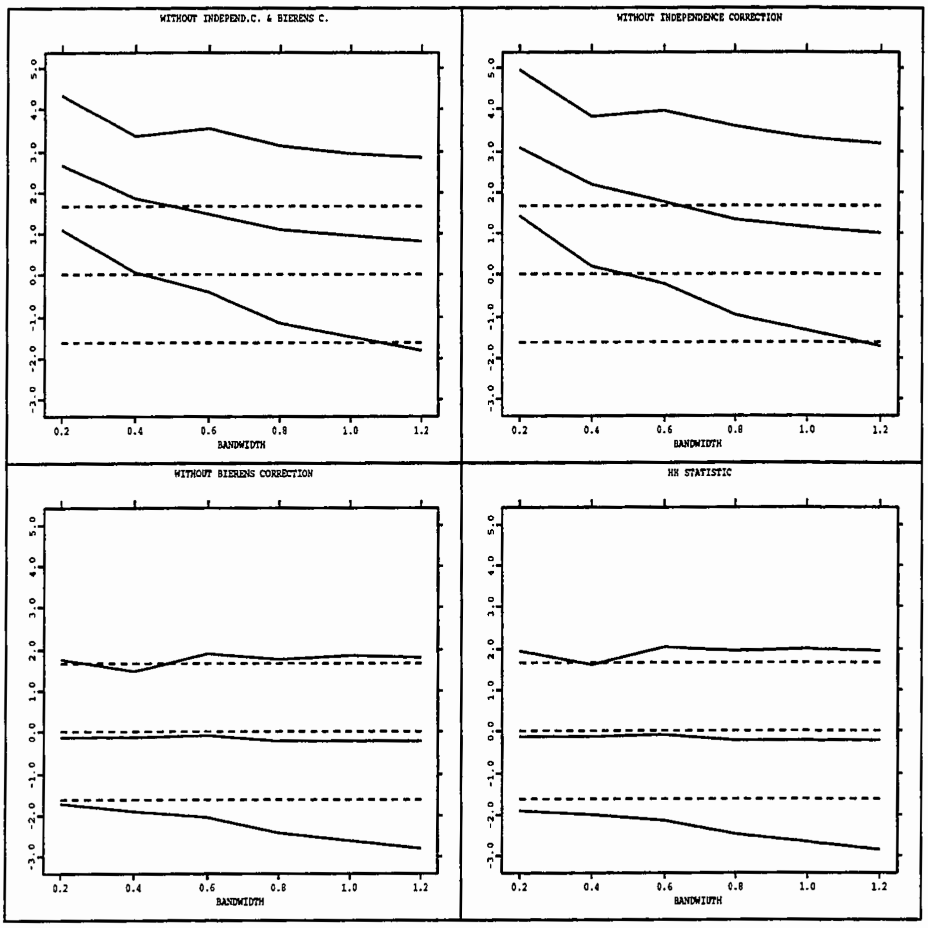

Figure 4.5: Effect of bias and independence corrections. 400 *observations, logit model, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

suggest a good performance of the statistic but the statistic was calculated under the ideal situation where the index function is completely known.

In practice to calculate the HH-test statistic the coefficients of the index function have to be estimated. The usual procedure estimates the index under the null by maximum likelihood which has the advantage of a much more easy implementation of the test than if the index was estimated under the semiparametric alternative.

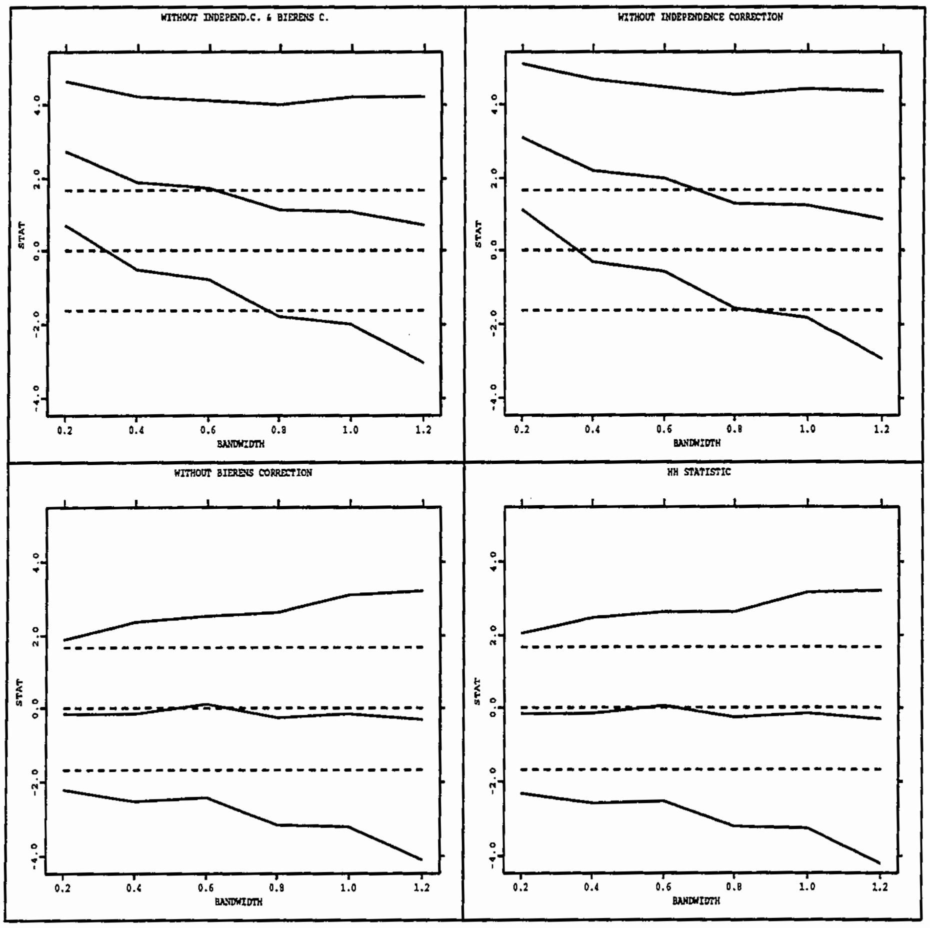

Figure 4.6: Effect of bias and independence corrections. 1000 *observations, logit model, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

If there is misspecification one may expect that the index estimate under the null will not be very accurate and this would have negative consequences on the power of the test. However, remembering the result of Ruud (1983) already mentioned in chapter 1, the parametric estimate for the slope parameters of the index can be consistent up to a factor scale if the link function is misspecified when the explanatories are normally distributed which is the case in this simulation study. Even under the null, the estimation of β introduces a certain "noise" into the statistic. Although this "noise" is asymptotically zero it

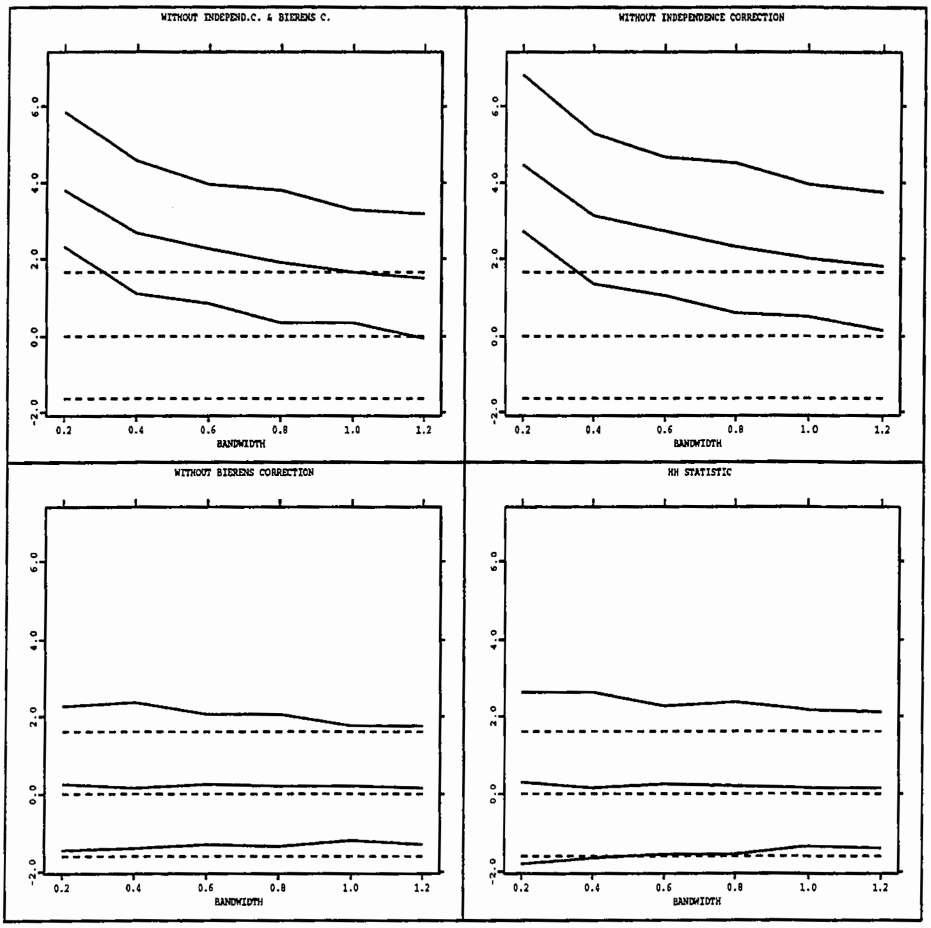

Figure 4.7: Effect of bias and independence corrections. 100 *observations, logit with a bump, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

can have a significant effect for samples of finite size damaging the appropriate behavior pointed out in section 4.3.

This section analyzes the behavior of the HH–test statistic by simulations when the coefficients of the index function are estimated by maximum likelihood. As before the designs considered are the logit and the logit with a bump corresponding respectively to (1.21) and (1.23). The HH–test statistic is computed according to (3.7) and the values of the statistic standardized by

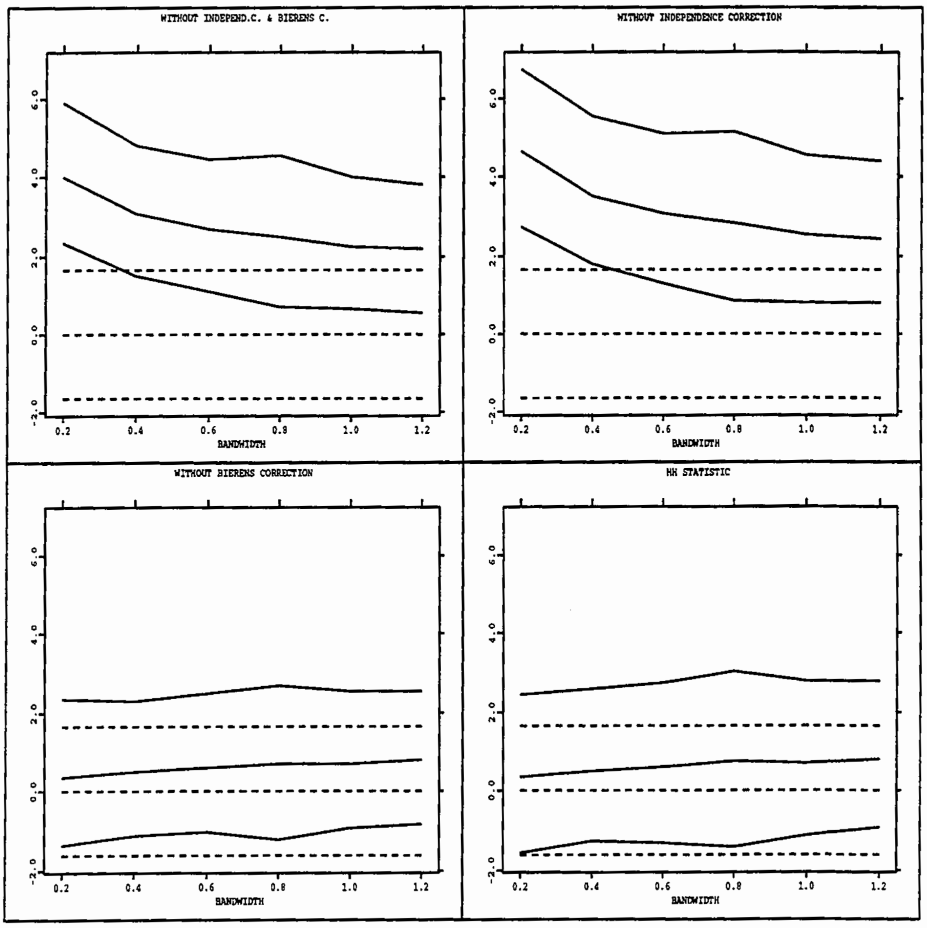

Figure 4.8: Effect of bias and independence corrections. 200 *observations, logit with a bump, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

the standard deviation estimated by (3.8) are reported. The results are given in Figure 4.11 and tables 4.1 and 4.2.

Figure 4.11 shows the mean and 5th and 95th percentiles of the statistic (for estimated index function) under the null (line), under the alternative (dotted line) and the same percentiles and mean of a standard normal distribution (dashed line). The plots in the Figure reveal that the estimation of the index function induces a negative bias in the statistic under the null. That is, the

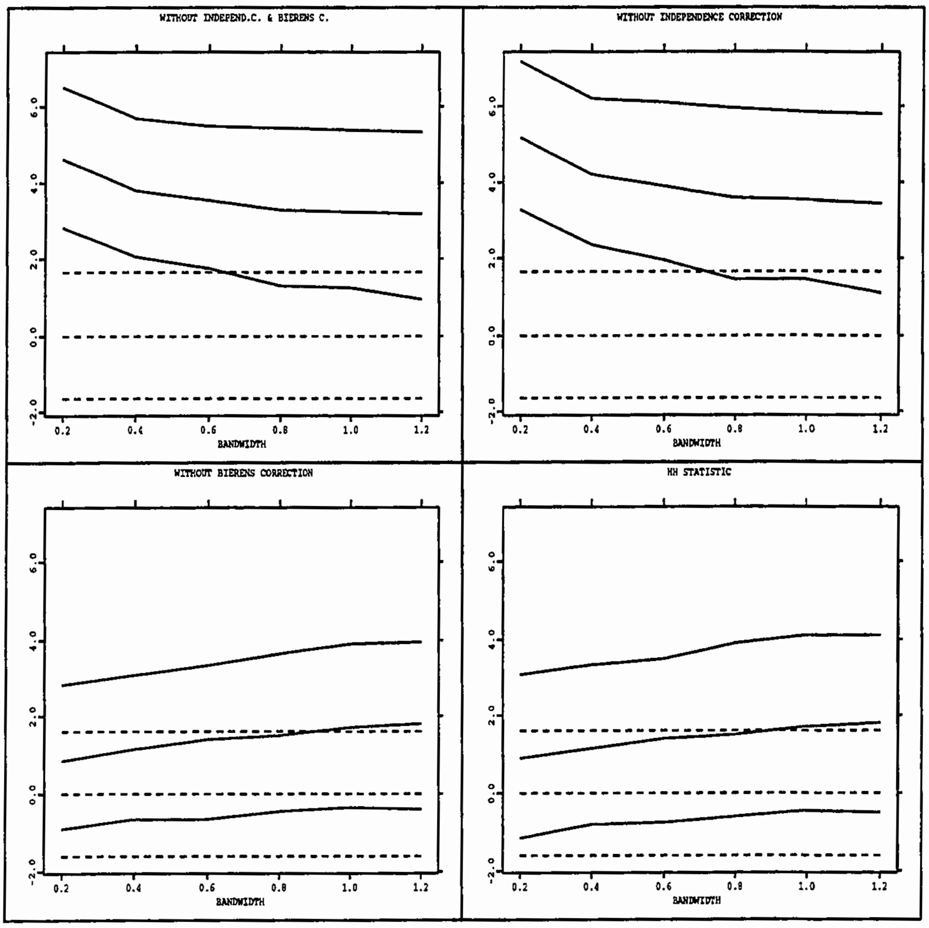

Figure 4.9: Effect of bias and independence corrections. 400 *observations, logit with a bump, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

empirical mean of the HH–test statistic is persistently below zero for all bandwidths even for samples with 1000 observations. One can also notice that the mean is closer to the 5th percentile than the 95th percentile suggesting that the statistic has a skewed distribution.

For small sample sizes the variance of the statistic starts to shrink for large bandwidths. However when the sample size gets larger the variance shrinks for the central values of the bandwidth and starts to expand for large bandwidths

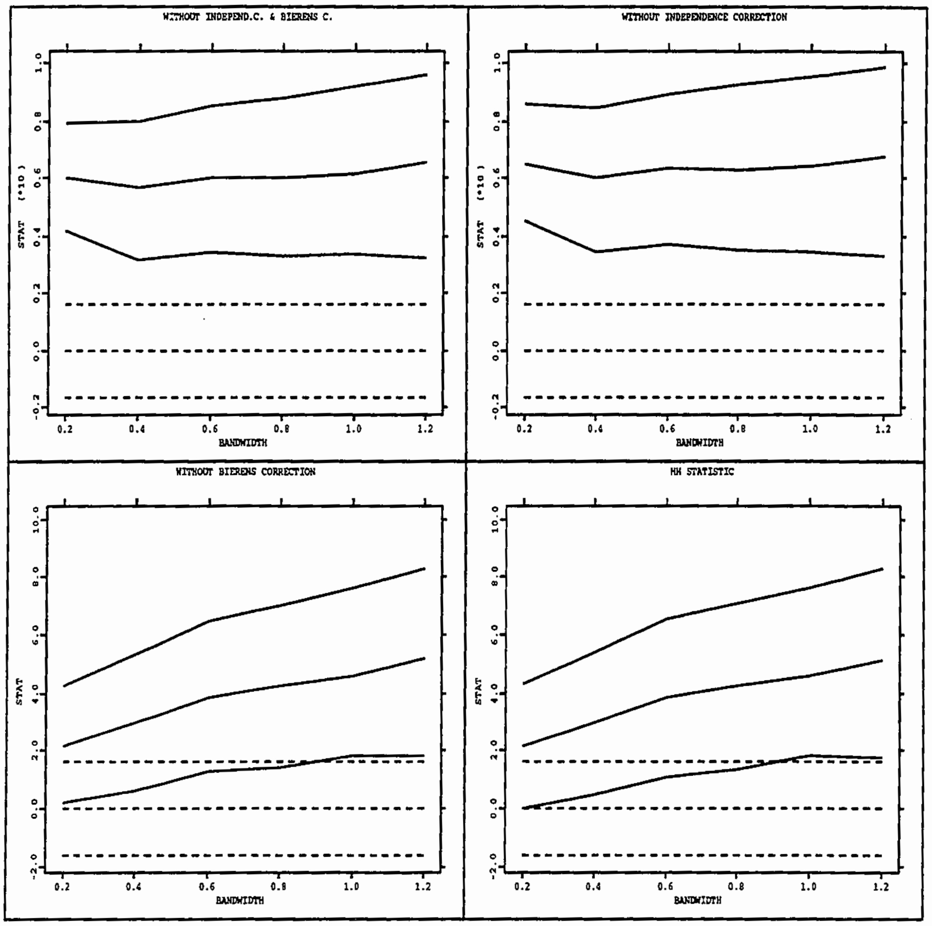

Figure 4.10: Effect of bias and independence corrections. 1000 *observations, logit with a bump, true index. Solid lines:* 5*th, mean and* 95*th empirical percentiles; dashed lines:* 5*th and* 95*th percentiles of standard normal distribution.*

(specially for $n = 1000$).

These results are more explicit in table 4.1. There it can be seen that the empirical mean of the statistic is negative and the median is even smaller; both remain almost the same for all sample sizes and decrease with increasing bandwidth. The 95th percentile and the 5th percentile also vary with the bandwidth reflecting the shrinking in the variance referred before followed by its expansion for the bigger sample sizes and large bandwidths.

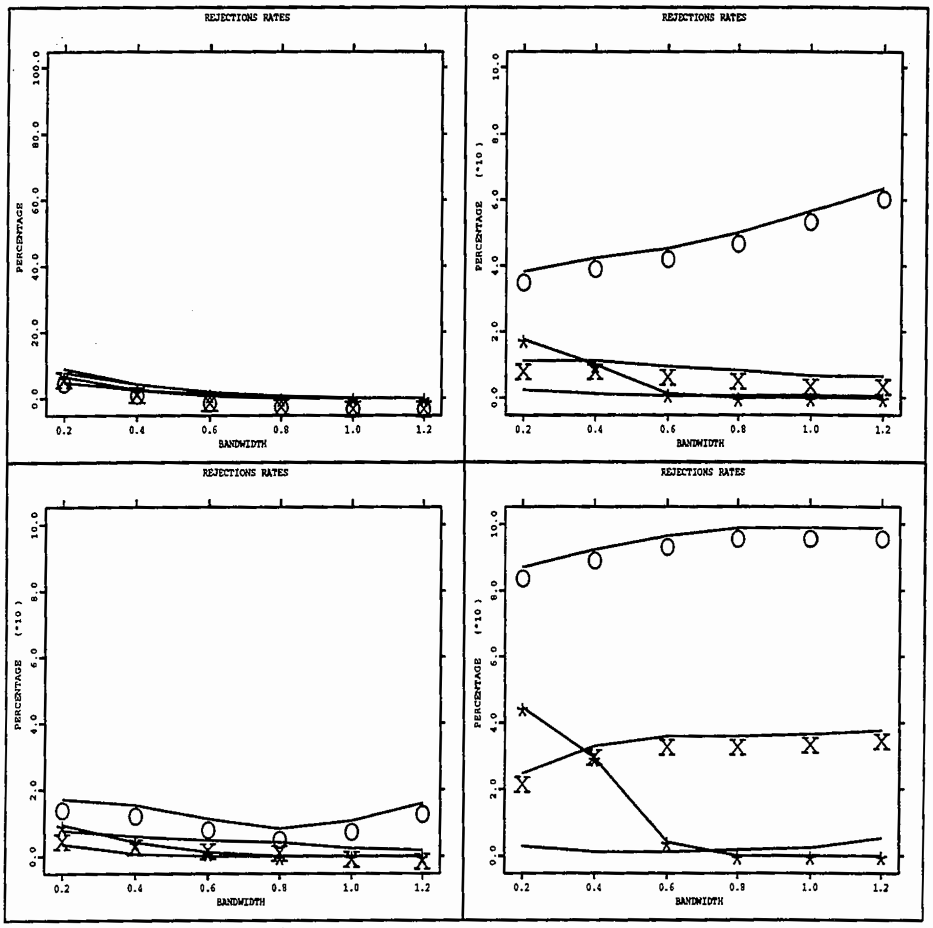

Figure 4.14: Rates of rejection of one–sided HH–test under the null (line) and under misspecification considering a logit with bump (line with circles), logit with heteroscedasticity (line with stars) and complementary log–log (line with crosses), for a nominal level of 10%.

for 200 observations the positive shift is less than 5% except for the logit with heteroscedasticity when $h > 0.8$. For the complementary log–log and the logit with bump the rejection rates decrease when h increases.

Note that the test is more able to detect the alternative logit with heteroscedasticity than the others when the bandwidth is set to $h = 1$ and $h = 1.2$ for samples with 100 and 200 observations. However, this ability to reject is

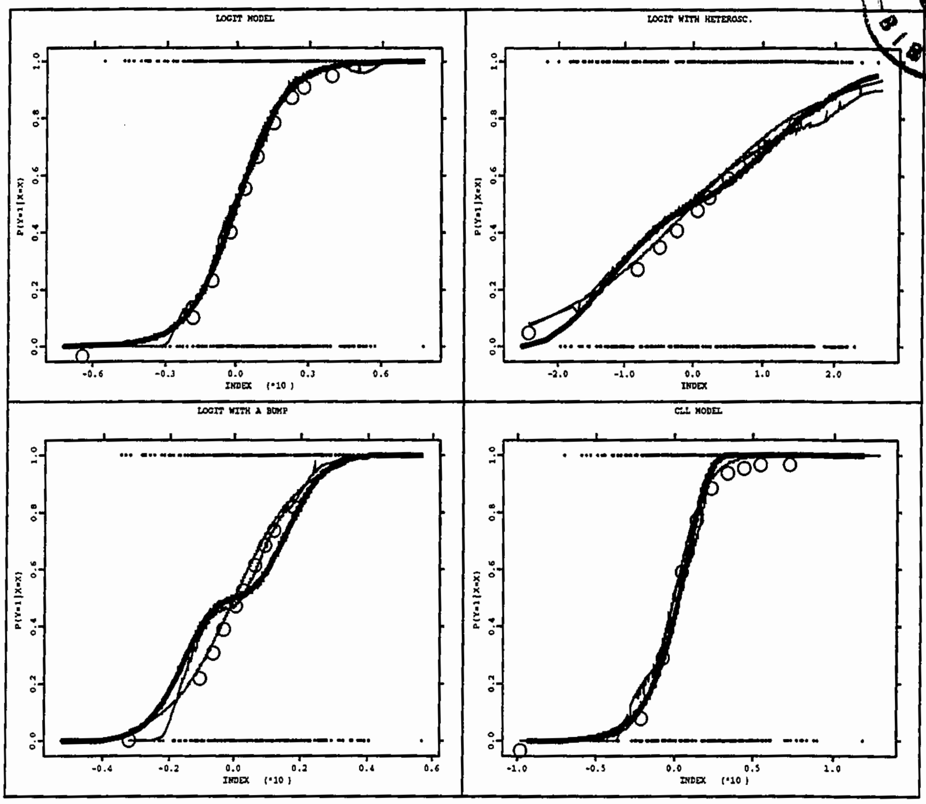

Figure 4.15: True model (thick line) with the kernel estimate (h = 0.4) and the parametric estimate assuming a logit (line with circles).

due to a severe oversmoothing in the kernel regression. The upper right plot in Figure 4.15 shows this oversmoothing for a sample with 400 points. On the other side, the test shows a severe lack of power to detect misspecification when data is generated from a logit with a bump or a complementary log–log for sample of size 200.

When the sample size grows to 400 and 1000 observations the empirical power is improved. However the rejection rates for the complementary log–log are far from being satisfactory even for 1000 observations staying below 28% as is patent in table 4.6. The logit with bump achieves better rejections rates when the bandwidth is large. For 400 observations table 4.5 shows that the best rate achieved is 50.4% for $h = 1.2$ at the 10% nominal level. However for $n = 1000$ the empirical power is greater than 80% for all bandwidths and greater than

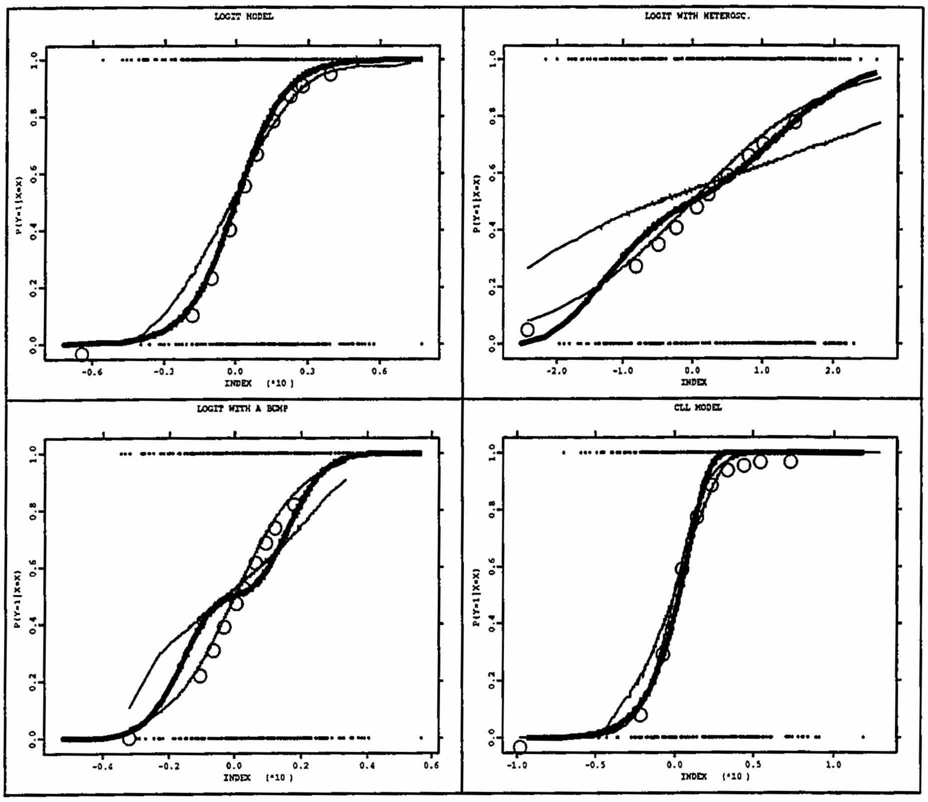

Figure 4.16: True model (thick line) with the kernel estimate (h = 1) and the parametric estimate assuming a logit (line with circles).

98% for $h = 1.2$. The empirical power for the logit with heteroscedasticity is very unsatisfactory for bandwidths smaller than 0.8. However for bandwidths $h = 1$ and $h = 1.2$, it yields approximately the same performance as for the logit with bump. For 1000 observations and $h = 1.2$ the rejection rate is about 95%.

The empirical size of the two–sided test is still much below than the nominal size for samples with 400 observations except when $h = 0.2$ or $h = 1.2$. For samples with 1000 observations the empirical size grows too quickly for the largest bandwidths. Table 4.6 shows that for $h = 1$ the empirical size is about 16% while for $h = 1.2$ is about 24% for a nominal size of 10%. This fact confirms the dependency under the null of the HH–test on the bandwidth chosen already noted before with the pernicious consequence of inducing the rejection of a true

	two–sided		one–sided		two–sided		one–sided	
	10%	5%	10%	5%	10%	5%	10%	5%
	Logit				Logit with a bump			
h = 0.2	7.6	2.8	4.6	2.4	7.6	2.8	7.4	4.2
	(1.19)	(0.74)	(0.94)	(0.68)	(1.19)	(0.74)	(1.17)	(0.90)
h = 0.4	0.8	0.2	2.2	0.4	2.8	1.8	4.4	2.8
	(0.40)	(0.20)	(0.66)	(0.28)	(0.74)	(0.59)	(0.92)	(0.74)
h = 0.6	0.2	0.0	0.8	0.2	0.8	0.6	1.8	0.8
	(0.20)		(0.40)	(0.20)	(0.40)	(0.35)	(0.59)	(0.40)
h = 0.8	0.0	0.0	0.2	0.0	0.0	0.0	0.6	0.0
			(0.20)				(0.35)	
h = 1.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
					(0.28)			
h = 1.2	0.4	0.0	0.0	0.0	1.0	0.2	0.0	0.0
	(0.28)				(0.44)	(0.20)		
	Logit with Heteros.				CLL			
h = 0.2	6.4	3.6	6.4	4.4	13.6	6.2	8.8	5.6
	(1.09)	(0.83)	(1.09)	(0.92)	(1.53)	(1.08)	(1.27)	(1.03)
h = 0.4	2.0	1.0	2.6	1.8	3.6	1.2	4.0	2.4
	(0.20)	(0.44)	(0.71)	(0.59)	(0.83)	(0.49)	(0.88)	(0.68)
h = 0.6	0.8	0.4	1.0	0.8	0.4	0.4	1.8	0.4
	(0.40)	(0.28)	(0.44)	(0.40)	(0.28)	(0.28)	(0.59)	(0.28)
h = 0.8	0.8	0.2	0.0	0.0	0.4	0.2	0.6	0.4
	(0.40)	(0.20)			(0.28)	(0.20)	(0.35)	(0.28)
h = 1.0	7.0	1.6	0.0	0.0	0.2	0.0	0.2	0.2
	(1.14)	(0.56)			(0.20)		(0.20)	(0.20)
h = 1.2	17.6	7.0	0.0	0.0	0.0	0.0	0.0	0.0
	(1.70)	(1.14)						

Table 4.3: Percentages of rejections of the HH–test for 100 *observations. Standard errors between brackets.*

null hypothesis.

The rejection rates in the one–sided test are in general greater than those in the two–sided test specially for the logit with a bump. The dependency on the bandwidth under the null is drastically reduced. Table 4.6 shows that the empirical size for samples with 1000 observations is smaller than 6% for a

	two–sided 10%	two–sided 5%	one–sided 10%	one–sided 5%	two–sided 10%	two–sided 5%	one–sided 10%	one–sided 5%
	Logit				Logit with bump			
h = 0.2	6.4	1.8	3.6	1.8	10.8	6.0	17.0	10.2
	(1.09)	(0.59)	(0.83)	(0.59)	(1.39)	(1.06)	(1.68)	(1.35)
h = 0.4	0.8	0.4	1.0	0.4	9.2	5.6	15.2	9.2
	(0.40)	(0.28)	(0.44)	(0.28)	(1.29)	(1.03)	(1.61)	(1.29)
h = 0.6	0.2	0.2	0.4	0.2	4.8	3.0	11.4	4.8
	(0.20)	(0.20)	(0.28)	(0.20)	(0.96)	(0.76)	(1.42)	(0.96)
h = 0.8	0.0	0.0	0.0	0.0	3.2	1.0	8.6	3.2
					(0.79)	(0.44)	(1.25)	(0.79)
h = 1.0	0.2	0.0	0.0	0.0	3.0	1.2	11.0	3.0
	(0.20)				(0.76)	(0.49)	(1.40)	(0.76)
h = 1.2	2.6	0.4	0.0	0.0	5.2	1.8	16.2	5.2
	(0.71)	(0.28)			(0.99)	(0.59)	(1.65)	(0.99)
	Logit with Heteros.				CLL			
h = 0.2	6.8	3.8	9.8	5.8	7.2	4.0	8.0	5.4
	(1.13)	(0.86)	(1.33)	(1.05)	(1.16)	(0.28)	(1.21)	(1.01)
h = 0.4	3.4	1.8	4.6	3.4	5.8	2.6	6.0	4.4
	(0.81)	(0.59)	(0.94)	(0.81)	(1.05)	(0.71)	(1.06)	(0.92)
h = 0.6	0.8	0.4	1.6	0.4	3.6	2.0	5.2	3.6
	(0.40)	(0.28)	(0.56)	(0.28)	(0.83)	(0.20)	(0.99)	(0.83)
h = 0.8	9.0	2.6	0.0	0.0	2.2	1.4	4.2	2.2
	(1.28)	(0.71)			(0.66)	(0.53)	(0.90)	(0.66)
h = 1.0	27.2	14.2	0.0	0.0	1.6	0.4	2.8	1.6
	(1.99)	(1.56)			(0.56)	(0.28)	(0.74)	(0.56)
h = 1.2	40.2	25.2	0.0	0.0	0.6	0.0	1.8	0.6
	(2.19)	(1.94)			(0.35)		(0.59)	(0.35)

Table 4.4: Percentages of rejections of the HH–test for 200 *observations. Standard errors between brackets.*

nominal size of 10%.

The empirical power for samples with 100 and 200 observations is still practically nonexistent. With the increase of the sample size to 400 and 1000 observations the empirical power is improved. Although, the complementary log–log has better rejections rates with the one–sided test the results are still very un-

	two–sided		one–sided		two–sided		one–sided	
	10%	5%	10%	5%	10%	5%	10%	5%
	Logit				Logit with a bump			
h = 0.2	6.6	1.6	2.6	1.6	31.4	25.2	38.2	31.2
	(1.11)	(0.56)	(0.71)	(0.56)	(2.08)	(1.94)	(2.17)	(2.07)
h = 0.4	1.6	0.4	1.0	0.6	35.0	27.6	42.4	35.0
	(0.56)	(0.28)	(0.44)	(0.35)	(2.13)	(2.00)	(2.21)	(2.13)
h = 0.6	0.4	0.4	0.6	0.4	34.4	27.8	45.4	34.4
	(0.28)	(0.28)	(0.35)	(0.28)	(2.12)	(2.00)	(2.23)	(2.12)
h = 0.8	1.0	0.0	0.4	0.2	36.2	28.4	49.8	36.2
	(0.44)		(0.28)	(0.20)	(2.15)	(2.02)	(2.24)	(2.15)
h = 1.0	2.4	0.8	0.6	0.0	44.2	32.2	56.2	44.2
	(0.68)	(0.40)	(0.35)		(2.22)	(2.09)	(2.22)	(2.22)
h = 1.2	8.6	3.6	0.8	0.2	50.4	42.2	63.6	50.4
	(1.25)	(0.83)	(0.40)	(0.20)	(2.24)	(2.21)	(2.15)	(2.24)
	Logit with Heteros.				CLL			
h = 0.2	13.8	8.2	17.8	11.8	9.2	5.0	11.2	7.6
	(1.54)	(1.23)	(1.71)	(1.44)	(1.29)	(0.97)	(1.41)	(1.19)
h = 0.4	6.0	3.6	10.2	6.0	7.8	4.8	11.2	7.6
	(1.06)	(0.83)	(1.35)	(1.06)	(1.20)	(0.96)	(1.41)	(1.19)
h = 0.6	3.2	0.8	1.0	0.8	6.4	3.8	9.4	6.4
	(0.79)	(0.40)	(0.44)	(0.40)	(1.09)	(0.86)	(1.31)	(1.09)
h = 0.8	31.6	17.4	0.0	0.0	4.8	3.0	8.0	4.8
	(2.08)	(1.70)			(0.96)	(0.76)	(1.21)	(0.96)
h = 1.0	58.2	44.4	0.0	0.0	4.0	2.4	6.4	4.0
	(2.21)	(2.22)			(0.28)	(0.68)	(1.09)	(0.28)
h = 1.2	67.6	59.0	0.0	0.0	3.6	1.8	6.4	3.0
	(2.09)	(2.20)			(0.83)	(0.59)	1.09)	(0.76)

Table 4.5: Percentages of rejections of the HH–test for 400 *observations. Standard errors between brackets.*

satisfactory given that even for $n = 1000$ the best rate achieved is about 38%. Curiously the one–sided test doesn't work at all for data generated from the logit with heteroscedasticity for $h > 0.4$. However, note that for $h = 0.2$ the rejection rate is superior than the rejection rates obtained for the CLL.

Figures 4.15 and 4.16 show the designs used in the simulation study (thick

	two–sided 10%	two–sided 5%	one–sided 10%	one–sided 5%	two–sided 10%	two–sided 5%	one–sided 10%	one–sided 5%
	Logit				Logit with a bump			
h = 0.2	5.8	1.8	3.2	2.0	82.2	74.8	86.6	74.8
	(1.05)	(0.59)	(0.79)	(0.20)	(1.71)	(1.94)	(1.52)	(1.94)
h = 0.4	1.2	0.6	1.6	1.0	89.8	86.0	92.4	89.8
	(0.49)	(0.35)	(0.56)	(0.44)	(1.35)	(1.55)	(1.19)	(1.35)
h = 0.6	0.8	0.4	1.6	0.8	94.2	90.4	96.2	94.2
	(0.40)	(0.28)	(0.56)	(0.40)	(1.05)	(1.32)	(0.86)	(1.05)
h = 0.8	6.0	0.8	1.8	1.2	96.4	93.8	98.6	96.4
	(1.06)	(0.40)	(0.59)	(0.49)	(0.83)	(1.08)	(0.53)	(0.83)
h = 1.0	15.8	9.0	2.6	1.4	97.8	96.2	98.6	97.8
	(1.63)	(1.28)	(0.71)	(0.53)	(0.66)	(0.86)	(0.53)	(0.66)
h = 1.2	23.8	16.4	5.4	2.8	98.2	97.4	98.6	98.2
	(1.90)	(1.66)	(1.01)	(0.74)	(0.59)	(0.71)	(0.53)	(0.59)
	Logit with Heteros.				CLL			
h = 0.2	38.8	30.8	45.0	38.8	17.6	13.4	24.8	13.4
	(2.18)	(2.06)	(2.22)	(2.18)	(1.70)	(1.52)	(1.93)	(1.52)
h = 0.4	22.8	17.6	29.8	22.8	23.0	17.2	32.8	23.0
	(1.88)	(1.70)	(2.05)	(1.88)	(1.88)	(1.69)	(2.10)	(1.88)
h = 0.6	16.6	8.8	4.6	3.2	27.2	18.2	35.8	27.2
	(1.66)	(1.27)	(0.94)	(0.79)	(1.99)	(1.73)	(2.14)	(1.99)
h = 0.8	73.0	63.2	0.0	0.0	27.4	21.0	36.0	27.4
	(1.99)	(2.17)			(1.99)	(1.82)	(2.15)	(1.99)
h = 1.0	91.4	86.8	0.0	0.0	27.8	20.6	36.4	27.6
	(1.25)	(1.51)			(2.00)	(1.81)	(2.15)	(2.00)
h = 1.2	94.6	92.2	0.0	0.0	27.8	21.0	37.8	27.6
	(1.01)	(1.20)			(2.00)	(1.82)	(2.17)	(2.00)

Table 4.6: Percentages of rejections of the HH–test for 1000 *observations. Standard errors between brackets.*

line) for the same sample as in Figure 4.1 together with the parametric estimate assuming a logit (line with circles) and the leave–one–out bias corrected kernel regression (line) given by (3.4) with bandwidth $h = 0.4$ (Figure 4.15) and bandwidth $h = 1$ (Figure 4.16). With correct specification (upper left plot) the parametric fit coincides with the true model and the kernel regression for $h = 0.4$ while for $h = 1$ the kernel regression deviates slightly from both in the upper tail and a little more on the lower tail. This deviation due to oversmoothing becomes more severe for larger bandwidths or larger sample sizes and explains the too big rate of rejections under the null pointed before. Note that the kernel regression is able to catch the bump in the logit with a bump and logit with heteroscedasticity for $h = 0.4$. When the bandwidth is set to 1 there is a severe oversmoothing in the mentioned models. On the other hand, the kernel regression and the logit regression are very close to the complementary log–log which clarifies about the unsatisfactory performance of the test for this model.

4.6 Concluding Remarks

This chapter studied the performance of the HH–test. It starts by analyzing the behavior of the statistic in "ideal conditions" that is, when the index function is known and has not to be estimated. In this situation the HH–test statistic has shown an appropriate behavior in the sense that the mean, standard deviation and the 5% and 10% percentiles were close to the respective quantities of the standard normal if the bandwidth was not too large. For this behavior the independence correction was essential while the Bierens' correction had a negligible effect.

Although the estimation of the coefficients in the index function by maximum likelihood (under the null) has asymptotically no effect, in finite samples it provokes a distortion of the distribution of the statistic. The distortion consists of a negative bias which persists even in samples with 1000 observations, a skewness of the distribution, and a shrinking or an expansion of the variance depending on the bandwidth and sample size.

The distortion mentioned above has some prejudicial consequences on the performance of the HH–test. Firstly, the distribution of the statistic under the null depends on the bandwidth for the two–sided test with the inconvenient that for large bandwidths the empirical size of the test becomes too large inducing false rejections. Secondly, the negative bias damages the power of the test when the asymptotical critical values from the standard normal are used.

The performance of the HH–test in rejecting a false model depends, as expected, on the particular alternative considered. Therefore, the HH–test has shown to be more able to detect the alternative logit with a bump. In particular for samples with 1000 observations the empirical power for this alternative is remarkably good. Remind that Figure 4.15 shows that the misspecified parametric logit fit deviates clearly from the logit with a bump while it is very close to the logit with heteroscedasticity and to the CLL (specially to the CLL) explaining the bad performance of the HH–test in these last two models.

The one–sided test is almost bandwidth robust under the null. However, the empirical size and empirical power are not much better than those obtained with the two–sided test. It has even an inferior performance than the two–sided test for great values of the bandwidth. However this superior performance of the later may be due merely to oversmoothing. In practical applications it can have a pernicious effect because it entails a too great first–type error, that is, a too great possibility of rejecting a well specified parametric model. The rejection rates of the logit for $h = 1.0$ and $h = 1.2$ in samples with 1000 observations confirm this idea.

Improvements are necessary to the implementation of the HH–test in small samples. This improvements can be done under two different perspectives. One is based on finding critical values for the test that are more accurate than the asymptotical given by the standard normal, that is, to find critical values that are closer to the true critical values of the test than the respective standard normal. The other is based on modifying the HH–test statistic to correct from bias and bandwidth dependency under the null in order to turn the asymptotical critical values more accurate, that is make the true critical values of the test closer to the respective standard normal. Those approaches will be subject of the next chapters.

Chapter 5

Bootstrapping the HH–test

5.1 Introduction

The study of the performance of the HH–test statistic included in chapter 4 reveals that the statistic has a negative bias and its distribution under the null in finite samples depends on the bandwidth in a way not captured by the asymptotic normal distribution. Thus, the HH–test test shows an unsatisfactory lack of (empirical) power to detect unspecified misspecification of the link function in binary response models.

In this chapter, a better approximation to the distribution of the HH–test statistic in finite samples is attempted to be achieved by the use of the bootstrap method. The aim is to find critical values for the test that are more accurate than those provided by the standard normal distribution and consequently to improve the empirical power of the test.

Hall (1992) gives evidence that studentization improves the performance of the bootstrap. He proves for some problems that with studentization the bootstrap approximation to the true distribution of the statistic is superior than the classical normal approximation given that it makes an error of order $O_p(n^{-1})$ while the last has an error of order $O(n^{-1/2})$. Hoping to profit from this superiority, the bootstrap method included in this chapter was implemented also with studentization.

To prove a superiority of the bootstrap over the standard normal approximation one needs tools like Edgeworth Expansions. This is the approach taken by Hall and Titterington (1989) to prove the superiority of Monte Carlo tests

1. Given a sample (x_i, y_i), $i = 1, ..., n$ calculate the ML estimate for β in the index function,$\hat{\beta}$, assuming the parametric model $F(x_i^T\beta)$. Here ML with bias reduction according to Firth (1992) is advised to be used.
   Calculate the fitted index $v_i = x_i^T\hat{\beta}$, $i = 1, \cdots, n$.
   Determine the statistic T_n given in equation (3.7) together with the respective estimated standard deviation, $\hat{\sigma}_T$ given by (3.8).
2. Generate a bootstrap sample with size n by drawing y_i^* from the Bernoulli distribution with parameter $F(x_i^T\hat{\beta})$, $i = 1, \cdots, n$. Estimate the coefficients β by performing one iteration in the M.L. algorithm of step 1 from starting point $\hat{\beta}$ with the bootstrap sample (y_i^*, x_i). Mosbach (1992) proves that this estimate is also $\sqrt{n}$–consistent (assuming the parametric model).
   Calculate the bootstrapped HH–test statistic T_n^* according to (5.12) together with the estimate of its standard deviation, $\hat{\sigma}_T^*$ obtained from (3.8) replacing $\hat{\beta}$ by $\hat{\beta}^*$. .
3. Repeat step 3 B times.
4. Calculate the bootstrap critical value c_α^* for a test of size α. Hall and Titterington (1989) give a formula for c_α^* in one-sided test based on the following. Define M to be

$$M = (B + 1)(1 - \alpha) \tag{5.13}$$

   Order the B values $T_n^*/\hat{\sigma}_T^*$ obtaining $t_{n[1]}^* \cdots t_{n[B]}^*$.
   Consider $c_\alpha^* = t_{n[M]}^*$.
5. Accept H_0 at a α level if

$$T_n/\hat{\sigma}_T < c_\alpha^*.$$

   Otherwise accept that the parametric model is misspecified.

Algorithm 5.1: Parametric bootstrap for binary responses.

defined by (1.22). When data are generated from a logit H_0 is true. When data are generated from the logit with bump or the CLL H_0 is false and the bootstrap test should lead to rejection. Data were generated according to the description in chapter 1, section 1.8. The shapes of these models can be seen

1. Generate a pseudo sample of size n.
   First, generate the explanatories x_i, $i = 1, \cdots, n$. Choose β and fix the link function $F(\bullet)$.
   Generate y_i from a Bernoulli distribution with parameter $F(x_i^T\beta)$, $i = 1, \cdots, n$.
2. Call steps 1 to 5 of the bootstrap algorithm 5.1.
3. Repeat steps 1 and 2 S times obtaining S values for the standardized statistic
   $$(T/\hat{\sigma}_T)_1, ..., (T/\hat{\sigma}_T)_S$$
   and S bootstrap critical values $c_\alpha^{*1}, ..., c_\alpha^{*S}$.
4. Determine the bootstrap empirical power according to,
   $$\#\{(T/\hat{\sigma}_T)_i > c_\alpha^{*i}, i = 1, ..., S\}.$$

Algorithm 5.2: ***The simulation procedure to calculate the empirical power of the bootstrap test.***

in Figure 4.1 of chapter 4.

All the simulations have been performed in GAUSS using $S = 500$ pseudo data sets with sample size $n = 200$ and $B = 199$ bootstrap replications of each data set. B was chosen according to (5.13) in order to have $M = 190$ for a one-sided test at 10% level.

In conducting the HH–test, the values for $h = 1$ and $h = 0.3$ have been chosen. This choice was made after a graphical analysis of the kernel regression estimate. The value $h = 1$ oversmooths a little bit the data for the logit design and logit with a bump while $h = 0.3$ undersmooths the data. Note that the aim is not to find the best semiparametric estimate from the data but to have a tool that detects deviations from the assumed link function.

The weight function $u(v)$ has been considered identically equal to one within an interval limited by the 95% and the 5% percentiles of the ordered fitted values of v.

The results obtained using standard normal critical values, the HH–test, and bootstrap critical values, the BHH–test, are shown in Table 5.1. Together with a two-sided test a one-sided test was also tried (e.g. using the 90% percentile to test at a 10% significance level). The one-sided test is motivated by the fact

	two-sided		one-sided	
	HH-test	BHH-test	HH-test	BHH-test
		Logit		
h = 0.3	4.4	10.6	2.4	7.4
h = 1.0	1.4	8.4	0.0	15.0
		Logit with a bump		
h = 0.3	9.0	23.4	12.0	35.4
h = 1.0	1.8	48.6	4.6	60.0
		CLL		
h = 0.3	5.6	11.2	5.8	11.4
h = 1.0	1.0	21.0	1.4	30.6

Table 5.1: *Percentages of rejections using normal critical values and bootstrap critical values. Nominal size is* 10%

that the HH–test statistic has mean bigger than zero under the alternative and consequently this test can achieve higher power.

The empirical size of the test is measured by the percentage of rejections of the logit model in the 500 trials. The empirical size of the test given by the normal critical values is considerably below the nominal size of 10% which is better approximated by the bootstrap test. This is clearly the consequence of the negative bias observed in the statistic. The empirical power of the classical HH–test (percentage of rejections for logit with a bump and for the CLL) is completely inexistent in this example which was expected given that the sample size is small. The bootstrap achieves remarkably better empirical power specially for bandwidth $h = 1$. For this bandwidth and despite the small size of the sample the rejection rate for the logit with bump can be considered very satisfactory. As expected one-sided test leads to an improvement of the power.

The results are in agreement with the studies of Proença (1993) and Proença and Ritter (1993) and show again the negative bias in the statistic. The statistic is slightly sensitive to the bandwidth under the null. For the type of alternatives used, undersmoothing induces a significant loss on the power of the test. These results show clearly that the bootstrap is beneficial in this problem.

The shape of the simulated density of the standardized HH–test statistic under H_0 together with the shapes of the bootstrap densities for each experi-

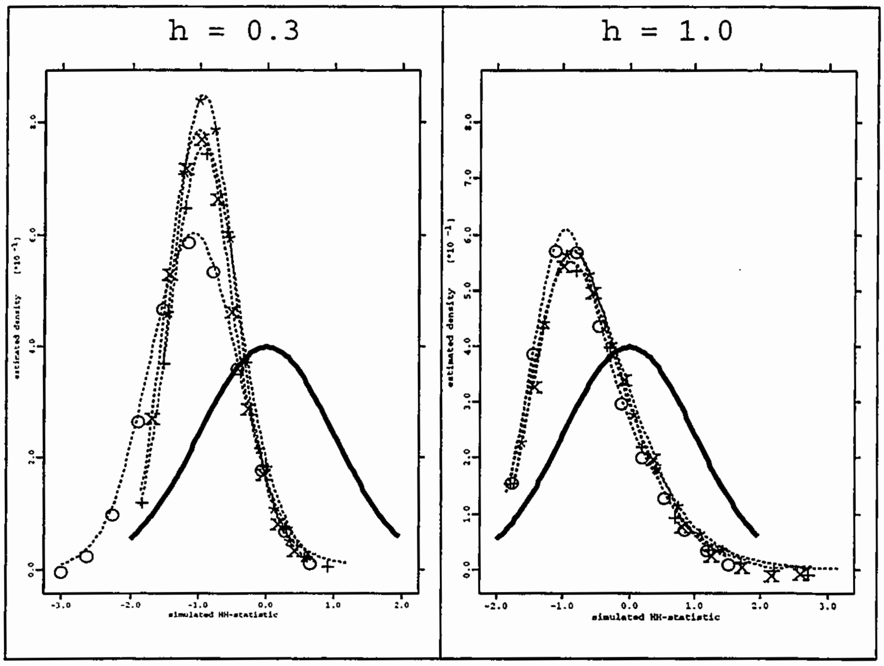

Figure 5.1: This Figure shows the simulated density of the HH–test statistic (plus), the bootstrap density under the null (cross), the bootstrap density under the alternative logit with bump (circle) and the bootstrap density under the alternative complementary log-log (star) for one sample (randomly chosen). The densities in the left window refer to the HH–test conducted with bandwidth equal to 0.3 while for the densities in the right window the HH–test was conducted with bandwidth equal to 1.

ment are presented in Figure 5.1. They correspond to one sample of the 500 pseudo-samples randomly chosen. The density estimates were obtained using kernel smoothing with a bandwidth of 0.8 and a Quartic kernel. The standard normal density is also included as a fat solid line in the same Figure. Again the negative bias and the bootstrap remedy are obvious.

The simulated density of the HH–test statistic is undoubtedly distinct from the standard normal having a clear mode left of zero. The bootstrap densities are very close to the simulated HH–test density whether under the null or the alternative (specially when the HH–test is conducted with $h = 1$) supporting the believe of accurate bootstrap critical values. For $h = 0.3$ the bootstrap density under the logit with a bump deviates a little from the HH–test simulated

	HH-stat	5% c.v.	90% c.v.	95% c.v.
	Travel mode-choice data			
h = 0.4	2.05	-1.33	0.39	0.70
h = 0.7	0.16	-1.69	0.34	0.68
	credit-scoring data			
h = 0.1	0.65	-1.55	0.59	1.20
h = 0.2	0.33	-1.38	0.65	1.06

Table 5.2: Results of the BHH–test on the mode-choice to travel and credit-scoring data sets. The HH–statistic divided by its estimated standard deviation and the bootstrap critical values for the two–sided and one-sided tests of size 10%.

density in the left tail and in kurtosis. However this deviation may not have consequences on defining inaccurate critical values specially for the one-sided test. The Figure reveals that the distribution of the HH–test statistic is bandwidth dependent and is negatively biased compared to the standard normal. Bootstrap is very successful on providing a better approximation to the true distribution of the HH–test statistic than the standard normal.

5.4 The Travel Mode Choice and Credit-Scoring Applications

In this section the bootstrap test is applied to the data sets under study in this dissertation: the travel mode-choice and credit-scoring.

Calculations were done with GAUSS according to algorithm 5.1. Given that bootstrap calculations are burdensome only two bandwidth values were tried for each data set. They were chosen based in the results of the HH–test included in section 3.3, chapter 3, with the aim to depict the different bandwidth–behavior of the statistic.

The results for the HH–test included in section 3.3 were performed with XploRe which uses the Quartic kernel while the program in GAUSS is defined for the Gaussian kernel. Given that the bandwidth depends on the particular kernel considered bandwidth values used here and in the mentioned section have to be different in order to obtain the same value for the HH–statistic. The procedure known as canonical transformantion described in Härdle (1990) allows to determine the bandwidth for the Gaussian kernel that gives approximately the same amount of smoothing as a given bandwidth used in the Quartic kernel. Therefore, the bandwidths in Table 5.2 were obtained by a canonical transformantion according to Härdle (1990) using XploRe 3. For the travel mode-choice problem the bandwidths $h = 0.4$ and $h = 0.7$ are equivalent respectively to bandwidths $h = 1.1$ and $h = 1.8$ for the Quartic kernel while for the credit-scoring problem bandwidths $h = 0.1$ and $h = 0.2$ are equivalent respectively to bandwidths $h = 0.3$ and $h = 0.5$ for the Quartic kernel.

The values assumed by the HH–statitic in both problems are consistent with the values in Table 3.1 and Table 3.2. For the travel-mode choice data the results of the BHH–test lead to the same conclusion as the HH–test in section 3.3. That is, for $h = 0.4$ the test indicates that the logit model should be rejected while it indicates that the logit is well specified if the bandwidth value is set to $h = 0.7$. However the bootstrap critical values are smaller than the respective critical values given by the standard normal, namely on the right tail, revealing the negative bias of the HH–statistic mentioned already in the precedent chapters.

For the credit-scoring problem the bootstrap one–sided test leads to rejection of the logit specification when $h = 0.1$ while the classical HH–test leads to acceptance. For $h = 0.2$ both HH–test and BHH–test give the same conclusion although the bootstrap critical values are smaller in magnitude than those given by the standard normal. Once more they reflect the negative bias of the HH–statistic.

To conclude, one may consider that the acceptance of the logit fit as result of the BHH–test when greater values of the bandwidth are used is due to a permanence in the power of the test of the pernicious effect of oversmoothing focused in section 3.3. On the other hand, the more accurate bootstrap critical values overcome the prejudicial effect of the negative bias in the HH–statistic by leading to rejection of the logit fit in the credit-scoring problem for $h = 0.1$ and the one–sided test while the HH–test indicates acceptance. Therefore, there are important forewarnings to consider a possible misspecification of the logit fit in both problems under study.

5.5 Concluding Remarks

A bootstrap procedure was proposed in this chapter with the aim to find more accurate critical values for the HH–test than the critical values provided by the standard normal approximation.

The bootstrap is perfomed under the assumed parametric model because the purpose is to determine a better approximation to the distribution of the HH–statistic under the null hypothesis than the one given by the standard normal.

The simulations performed have shown that the bootstrap test, **BHH**–test, achieves better rejection rates than the HH–test. On the other hand, the simulated bootstrap distribution shows to be closer to the simulated distribution of the HH–statistic under the null than the standard normal. Therefore, one may conclude that bootstrap reveals to be beneficial to improve the performance of the HH–test reducing the pernicious effect of the negative bias of the HH–statistic.

The bootstrap critical values obtained for the applications travel-mode choice and credit-scoring under study in this dissertation reveal the existence of the negative bias of the HH–statistic and indicate misspecification of the logit fit in the credit-scoring data while the HH–test leads to acceptance of the null.

Chapter 6

A modified HH–test

The analysis of the performance of the HH–test carried out in chapter 4 led to the conclusion that the HH–test behaves poorly in samples of small size when the asymptotical critical values of the standard normal are used. This poor behavior, expressed in a lack of power of the test, is due mainly to the presence of a persistent negative bias in the statistic (even in samples with 1000 observations) and to a bandwidth dependency of the variance of the statistic. These distortions are manifested when the index function is estimated by maximum likelihood under the null hypothesis and they are not present when the statistic is evaluated at the true index function. However, in practice the coefficients in the index function are unknown and have to be estimated. Maximum likelihood estimation under the null has the advantage of being easily implemented simplifying considerably the application of the test.

Chapter 5 presents a method to improve the performance of the HH–test in finite samples based on the bootstrap. The aim is to find more accurate critical values for the test than the asymptotical standard normal.

Another perspective to improve the performance of the test is based on the definition of analytical corrections to the HH–test statistic in order to eliminate the negative bias and to find an estimate of the variance in finite samples less dependent on the bandwidth. That is, the modifications would turn the asymptotical critical values more accurate. This is the approach taken in this chapter.

This chapter presents a modification to the HH–test statistic that aims to reduce its bandwidth dependency. This modification was proposed by Proença and Ritter (1994). The modified HH–test (from now on referred to as MHH–

not be explicitly derived although the expressions for the first and second moments corrected from the effect of using the residuals of the parametric fit were obtained but are complex and not easy to implement.

One advantage of the modified statistic relatively to the HH–test is that the asymptotic bias on the kernel regression is no longer present. In fact, $\tilde{r}(x_i^T\beta)$ has expectation zero conditional on x_i, $i = 1, \cdots, n$. Consequently Bierens' correction is not needed anymore and calculations become simpler. Another important aspect of this feature is related to the bandwidth. Now the bandwidth does not need to converge to zero at the rate $n^{-1/5}$ as in Horowitz and Härdle (1994) because there is no need to balance bias and variance in the mean integrated squared error. For the modified statistic the bandwidth can converge slower to zero which improves the power of the test under local alternatives to a level closer to parametric tests. The conclusion is that the modified statistic shows a stronger bandwidth independence under the null than the HH–test statistic. Thus, the problem of optimal bandwidth choice under the null is no longer relevant in this case.

Define $\tilde{r}$ as the vector containing the smoothed residuals $\tilde{r}_1, \cdots, \tilde{r}_n$ with $\tilde{r}_i = \tilde{r}(x_i^T\beta)$. $\tilde{r}$ can be written as $\tilde{r} = W\ r$, the product between the matrix that contains the smoothing weights (usually known as the smoothing matrix) and the vector of residuals $r = (r_1, \cdots, r_n)'$, where $r_i = r(x_i^T\beta)$. For the leave–one–out Nadaraya–Watson estimator the smoothing matrix has elements w_{ij} defined as

$$w_{ij} = \begin{cases} \frac{K[(x_j^T\beta - x_i^T\beta)/h]}{\sum\limits_{j\neq i} K[(x_j^T\beta - x_i^T\beta)/h]} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{6.2}$$

The MHH–test statistic can therefore be expressed as

$$MT_n = \sqrt{h}\, r^T U W\ r,$$

where the matrix U is a diagonal matrix with elements $u_i = u(x_i^T\beta)$. The matrix UW is not symmetric. In the following derivations, it is more convenient to work with a symmetrized version given by $W^* = (1/2)(UW + W^TU)$ which enables us to write MT_n as the quadratic form

$$MT_n = \sqrt{h}\, r^T W^*\ r.$$

Since the $r(x_i^T\beta)$ are mutually independent, so that $E[r_ir_j|x_ix_j] = 0$, have expectation conditional on x_i equal to zero under the null, and since the elements in the diagonal of W^* are zero the expectation of MT_n conditional on x_i is zero under the null. Straightforward calculations show that its variance is given by

$$\sigma_{MT}^2 = 2h \sum_i \sum_{j \neq i} \sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta){w_{ij}^*}^2$$

$$\sigma_{MT}^2 = h \sum_i \sum_{j \neq i} \sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta)(u(x_i^T\beta)w_{ij}^2 + u(x_i^T\beta)w_{ij}u(x_j^T\beta)w_{ji}).$$

with $\sigma^2(x_i^T\beta) = F(x_i^T\beta)\{1 - F(x_i^T\beta)\}$.

The modified version has the same asymptotic distribution under the null as the HH–test statistic as it is stated in the next theorem.

Theorem 6.1 *Under H_0 and under the assumptions in Horowitz and Härdle (1994) the MHH–test statistic is asymptotically distributed as $N(0, \sigma_T^2)$ with σ_T^2 as in equation (3.4).*

The proof of this theorem follows the proof in Horowitz and Härdle (1994) and is given in the appendix.

6.2 Effect of Estimating the Index Function

In practice, the residuals $r(x_i^T\beta)$ have to be substituted by the fitted residuals,

$$\hat{r}_i = r(x_i^T\hat{\beta}) = y_i - F(x_i^T\hat{\beta}).$$

where as before $\hat{\beta}$ is the maximum likelihood estimate under the null. The MHH–test statistic becomes,

$$MT_n = \sqrt{h}\,\hat{r}^T W^*\,\hat{r}, \tag{6.3}$$

where $\hat{r}$ is the vector with the n residuals $\hat{r}_1, \cdots, \hat{r}_n$ defined above and W^* is obtained as defined before but with β replaced by $\hat{\beta}$ in (6.2), the smoothing matrix.

An imediate consequence of estimating the coefficients in the index function is that the residuals $\hat{r}_i = r(x_i^T\hat{\beta})$ are no longer mutually independent. Therefore, the resulting statistic has no longer zero mean.

The following analyzes the effect of the estimation of the coefficients in the index function on the moments of the statistic when the parametric model is the logistic link. It makes use of the approximation of $\hat{r}_i$ to r_i for logistic regression employed in le Cessie and van Houwelingen (1991). Following these authors, the vector of fited residuals $\hat{r} = (\hat{r}_1 \cdots \hat{r}_n)'$, where $\hat{r}_i = r(x_i^T\hat{\beta})$, that is the vector of residuals of the parametric fit is related to the vector of "true" by

$$\hat{r} = (I - H)\, r + o_p(n^{-1/2}) \tag{6.4}$$

where H is the matrix

$$H = VX(X^T VX)^{-1}X^T$$

X is the design matrix with rows x_i, $i = 1, ..., n$ and V is a diagonal matrix with the weights $v_i = \sigma^2(x_i^T \beta)$. The expression for $\hat{r}$ in (6.4) results from a Taylor expansion of $y_i - F(x_i^T \hat{\beta})$ about its value for $\beta = \beta$ when $F(\bullet)$ is the logit link.

Given (6.4) the MHH–test statistic defined in (6.3) can be approximated by

$$AMT_n = \sqrt{h}\, r^T(I - H)^T W^*(I - H)\, r = \sqrt{h}\, r^T D\, r.$$

with $D = (I - H)^T W^*(I - H)$. The expectation of AMT_n conditional on X is given by

$$\begin{aligned} E[AMT_n] &= \sqrt{h}\, \sum_i d_{ii}\sigma^2(x_i^T \beta) & (6.5)\\ &= -tr(V^{1/2}W^*V^{1/2}H^*), & (6.6) \end{aligned}$$

where $H^* = V^{1/2}X(X'VX)^{-1}X'V^{1/2}$. In general, this expectation is nonzero and Proença and Ritter (1994) conjecture that it is negative in most cases. Note that W^* is symmetric, but not always positive definite. Nevertheless, the expression $tr(V^{1/2}W^*V^{1/2}H^*)$ can be rewritten as $tr[(X'VW^*VX)(X'VX)^{-1}]$ and one sees that it suffices that the trace of matrix $X^{*\prime}W^*X^*$ with the "standardized" design

$$X^* = VX(X'VX)^{-1/2}$$

is positive to make the expectation negative. This was the case in all situations studied by Proença and Ritter (1994) and in all examples in this chapter.

The variance of AMT_n conditional on x_i is equal to

$$\begin{aligned} \sigma^2_{AMT} &= h\sum_i d_{ii}^2\{E[r(x_i^T\beta)^4] - [\sigma^2(x_i^T\beta)]^2\} + 2\sum_i\sum_{j\neq i}\sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta)d_{ij}^2\\ \sigma^2_{AMT} &= h\sum_i d_{ii}^2\{\sigma^2(x_i^T\beta) - 4[\sigma^2(x_i^T\beta)]^2\} + 2\sum_i\sum_{j\neq i}\sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta)d_{ij}^2\\ &= 2tr(DVDV) + \sum_{i=1}^n d_{ii}^2[\sigma^2(x_i^T\beta) - 6\sigma^4(x_i^T\beta)]. \qquad (6.7) \end{aligned}$$

The conditional expectation and variance of AMT_n can be taken as approximations of the conditional expectation and variance of the MHH–test statistic evaluated for the fitted index function. They can be easily estimated under the null because the only unknowns are the conditional variances of the residuals $\sigma^2(x_i^T\beta)$. Therefore expression (6.6) with $\sigma^2(\bullet)$ substituted by $\hat\sigma^2(\bullet)$ gives the bias correction and expression (6.7) with $\sigma^2(\bullet)$ substituted by $\hat\sigma^2(\bullet)$ gives a bandwidth robust estimate of the variance of the statistic in finite samples.

Algorithm 6.1 calculates the MHH–test. This algorithm is implemented in XploRe 3 on the procedure `MODHHTST` which is included in the Appendix.

6.3 Alternative Bias Correction

To determine the bias correction presented in last section it is necessary to calculate the smoothing matrix W and the matrix H. Both are matrices of dimension $n \times n$. For small to moderate sample sizes they are not hard to compute. Also, Proença and Ritter (1994) propose an alternative way to determine the bias correction which does not use the matrices mentioned before. Moreover, this approach is more general and seems to have an extended applicability.

The bias in the HH–test and MHH–test statistics is caused by the dependence among the fitted residuals introduced by the estimation of the parameter vector β. A first order Taylor expansion of $r(x_i^T\hat\beta) = y_i - F(x_i^T\hat\beta)$ about its value in $r(x_i^T\beta) = y_i - F(x_i^T\beta)$, the residual for the assumed parametric model, gives

$$r(x_i^T\hat\beta) = r(x_i^T\beta) - (\hat\beta - \beta)x_i\sigma^2(x_i^T\beta) + o_p(n^{-1/2})$$

On the other hand $\hat\beta$ is a linear estimate of Y and thus can be written as $\hat\beta = C + AX^TY$. Substituting $\hat\beta$ in the above formula by this linear expression, after some trivial algebra one can write the following expectation

$$\begin{aligned}
E[r(x_i^T\hat\beta)r(x_j^T\hat\beta)] &= -2\sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta)x_i^TAx_j \\
&+ \sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta)x_i^TAX^TVXAx_j + o_p(1).
\end{aligned}$$

Suppose now that each fitted residual $r(x_i^T\hat\beta)$ is obtained without the i'th observation. That is, $r(x_i^T\hat\beta)$ is calculated using the fitted coefficients $\hat\beta_{-i} = C + AX_{-i}^Ty_{-i}$ where X_{-i} and y_{-i} denote the X matrix and the Y vector without row i respectively. The above formula yields

$$E[r(x_i^T\hat\beta)r(x_j^T\hat\beta)] = \sigma^2(x_i^T\beta)\sigma^2(x_j^T\beta)x_i^TAX_{-i}^TV_0X_{-j}Ax_j + o_p(1).$$

with V_0 a $n-1$ diagonal matrix with elements $v_1, \cdots, v_{i-1}, 0, \cdots, 0, v_{j+1}, \cdots, v_n$.

1. Estimate parametrically β in the index function, obtaining $\hat{\beta}$.
   Calculate the fitted index $\hat{v}_i = x_i^T \hat{\beta}$, $i = 1, \cdots, n$.
   Calculate the parametric residuals $\hat{r}_i = y_i - F(\hat{v}_i)$, $i = 1, \cdots, n$ and define the vector $\hat{r} = [r_1, \cdots, r_n]$.

2. Define the weight function $u(\hat{v}_i)$ to be equal to one for the 90% or 95% central values of the ordered $\hat{v}_i$, $i = 1, \cdots, n$.

3. Calculate the smoothing matrix W with elements w_{ij}, $i = 1, \cdots, n$ and $j = 1, \cdots, n$ where
$$w_{ij} = \begin{cases} \frac{K[(\hat{v}_j - \hat{v}_i)/h]u(\hat{v}_i)}{\sum\limits_{j \neq i} K[(\hat{v}_j - \hat{v}_i)/h]} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

4. Calculate the MHH–statistic
$$T_n = \sqrt{h} \sum_{i=1}^{n} \hat{r}^T W \hat{r}.$$

5. Calculate the estimated bias correction and variance:
   5.1 calculate the H matrix $H = \widehat{V}X(X^T\widehat{V}X)^{-1}X^T$ where X is the design matrix and $\widehat{V}$ is a diagonal matrix with the parametrically estimated conditional variances of Y_i, $\sigma^2(\hat{v}_i)$.
   5.2 calculate the matrix $D = (I - H)^T W (I - H)$
   5.3 calculate the bias correction according to
$$bc = \sqrt{h}\ \textstyle\sum_i d_{ii}\sigma^2(\hat{v}_i)$$
   5.4 calculate the estimate of the corrected variance
$$\hat{\sigma}_T^2 = 2tr(DVDV) + \sum_{i=1}^{n} d_{ii}^2[\sigma^2(\hat{v}_i) - 6\sigma^4(\hat{v}_i)].$$

6. Choose the size of the test to be α. Set $c_{\alpha/2}$ and c_α equal respectively to the quantiles $1 - \alpha/2$ and $1 - \alpha$ of a standard normal.
   Accept the parametric model if

   - $|(T_n - bc)/\hat{\sigma}_T| < c_{\alpha/2}$ for a two sided-test
   - $(T_n - bc)/\hat{\sigma}_T < c_\alpha$ for a one-sided test

   Otherwise, accept that the parametric model is misspecified.

Algorithm 6.1: The MHH–test

This correction reduces the covariance between the residuals since the first term in the preceding expression has vanished and the second is of smaller size. In practice this conduct amounts to calculate the statistic using the "leave–one–out" residuals $y_i - F(x_i^T \hat{\beta}_i)$ instead of the usual residuals $y_i - F(x_i^T \hat{\beta})$. The applied studies where this method was implemented suggest that it is as effective as the direct adjustment.

6.4 Performance of the Modified Statistic

This section analyzes the performance of the MHH–test statistic by a simulation study. For sake of comparison the simulations here were designed to mimic the simulations about the performance of the HH–test. Thus, the models underlying the data generation are the four different designs introduced in chapter 4, respectively the logit in (1.21), the logit with a bump given by (1.23), logit with heteroscedasticity defined by (1.24) and complementary log–log given by (1.22). The data were generated according to the description in section 1.8). The weight function is the same as before being defined in the mentioned section. The test was implemented for the same grid of bandwidths as in chapter 4, i.e. $0.2, 0.4, \cdots, 1.2$. The samples were generated with size 100, 200 and 400. Samples with 1000 observations were not tried here because they demanded a computational effort too big too afford. Anyway the features analyzed in this study are more compelling for samples of small size. For each study the number of simulations was set to 200. The number of simulations used in chapter 4, 500, was not possible to use here because computations are much more demanding. Note that an inferior number of simulations will produce a lower precision of the results.

This study will not be as detailed as the sudy in chapter 4. One reason is because computations are much more burdensome for the MHH–test. On the other side, here the aim is more focused on assessing the performance of the MHH–test in correcting from bias and bandwidth dependency of the variance and the respective consequences on the improvement of the empirical size and power relatively to the HH–test.

The study proceeds as follows. After generating a sample under the null (from a logit model) or under one of the three other alternative models considered (logit with bump, logit with heteroscedasticity and complementary log–log), the coefficients in the index function are estimated by maximum likelihood assuming a logit model obtaining the residuals $\hat{r}_i = y_i - F(x_i^T \hat{\beta})$. The MHH–test statistic is calculated according to (6.3). The explicit bias correction

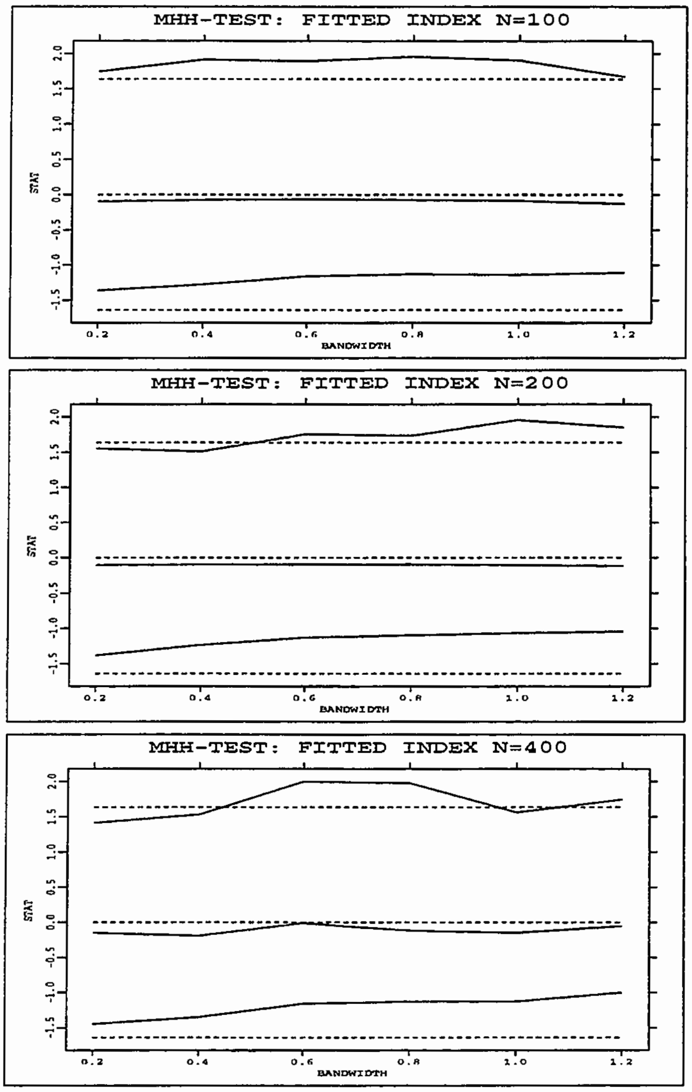

Figure 6.1: *MHH-test statistic with fitted index. Solid lines: 5th, mean and 95th empirical percentiles under H_0 – Logit link; dashed lines: 5th and 95th percentiles of standard normal distribution.*

defined in (6.6) is calculated together with the estimated standadard deviation according to (6.7). Then, the values of the statistic after subtraction from the bias correction and division by the estimated standard deviation are retained.

In a first step some descriptive statistics about the bias corrected standardized statistic under the null will be analyzed in a similar way as in section 4.4. Then after the study will concentrate on the assessment of the empirical size and power of the MHH–test.

Figure 6.1 shows the 5%th, 95%th empirical percentiles and the mean of the bias corrected standardized statistic together with the respective quantities of the standard normal. Now the picture is much more stable than for the HH–test.

Although the mean of the statistic is still below zero the negative bias is clearly smaller in absolute value than the bias in the HH–test. Table 6.1 shows that the bias lies between -0.01 and -0.19 while for the HH–test was between -0.34 and -0.76 (see table 4.1). The test shows some weekness on approximating the 5%th percentile of the standard normal. The respective percentile of the statistic is systematically greater than the latter and shows an increasing tendency with the increase of the bandwidth. The variance of the MHH–test statistic can be considered satisfactorily stable. It is always very close to one varying between 0.89 and 1.02.

Figure 6.2 shows the simulated densities of the MHH–test statistic obtained by kernel smoothing with bandwidth 0.8 and the standard normal density (thick line). The different densities result from cconducting the test with respectively $h = 0.2$ (line with circles), $h = 0.4$ (line), $h = 0.6$ (line with stars), $h = 0.8$ (line with crosses), $h = 1$ (line with plus) and $h = 1.2$ (dotted line). The mode of the MHH–test distribution is negative but close to zero and the distribution is slightly skweed to the right. There is still some bandwidth dependency on the shape of the densities but much less than in the case of the HH–test.

The results in figures 6.1 and 6.2 and in table 6.1 allow to conclude that the MHH–test has achieved satisfactorily the two main goals proposed. The corrections seem to be effective in reducing the bias and almost eliminating the bandwidth dependency of the distribution of the statistic under the null. The robustness to the bandwidth is translated by a stable estimate of the standard deviation very close to the asymptotic one. These results are not completely meaningful if not associated with an improved power of the test. This will be the subject of investigation in the following.

The results about empirical power and empirical size of the MHH–test statistic are reported in figures 6.3 and 6.4 and tables 6.2, 6.3 and 6.4.

The empirical size of the two–sided test is measured by the percentage of the 200 simulations run for which the absolute value of the bias corrected standardized MHH–test statistic was bigger than the asymptotical critical value

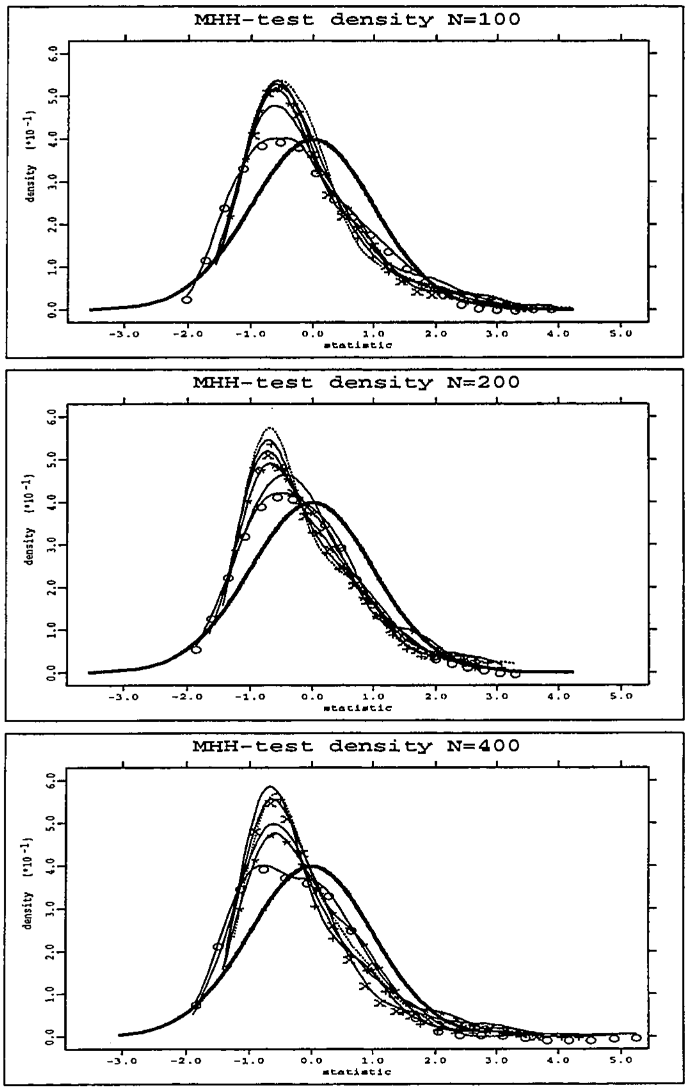

Figure 6.2: Simulated densities of the MHH-test statistic under H_0 conducted with $h = 0.2$ (line with circles), $h = 0.4$ (line), $h = 0.6$ (line with stars), $h = 0.8$ (line with crosses), $h = 1$ (line with plus), $h = 1.2$ (dotted line) and standard normal (thick line).

(that is the percentage of rejections or rejection rate) when the data is generated from a logit model, design (1.21). The empirical power is given by the percentage of rejections when the data is drawn from the other designs used

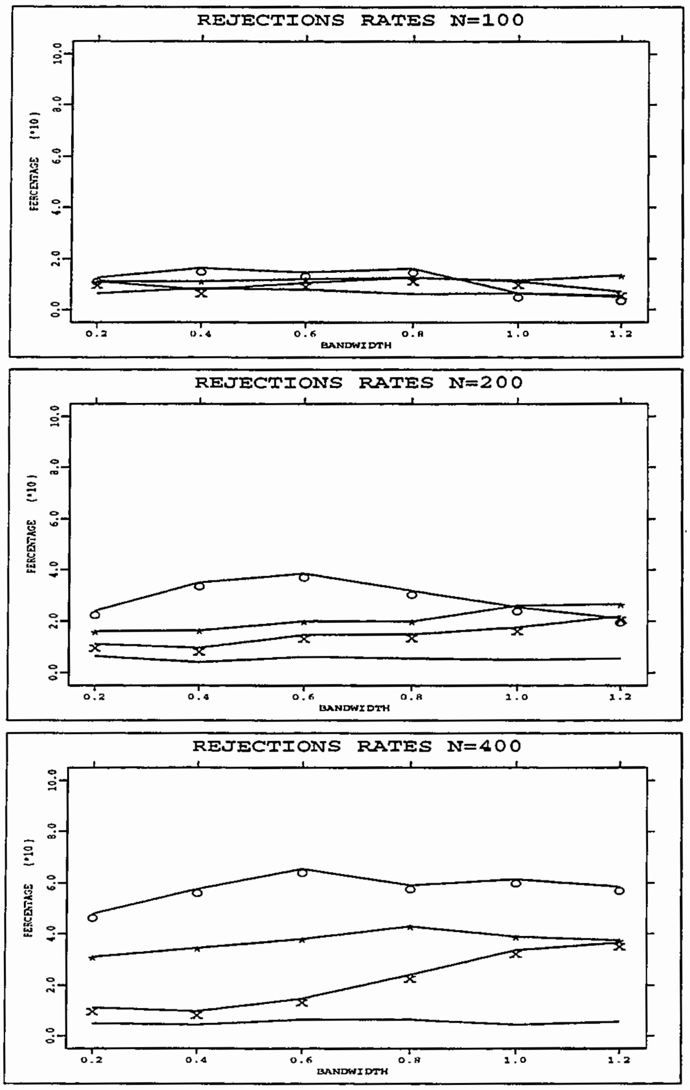

***Figure 6.3:** Rates of rejection of two–sided MHH–test under the null (line) and under misspecification considering a logit with bump (line with circles), logit with heteroscedasticity (line with stars) and complementary log–log (line with crosses), for a nominal level of* 10%.

but it is estimated parametrically assuming a logit. At a nominal level of 10% and 5% the critical values of the two–sided test are respectively 1.645 and 1.96.

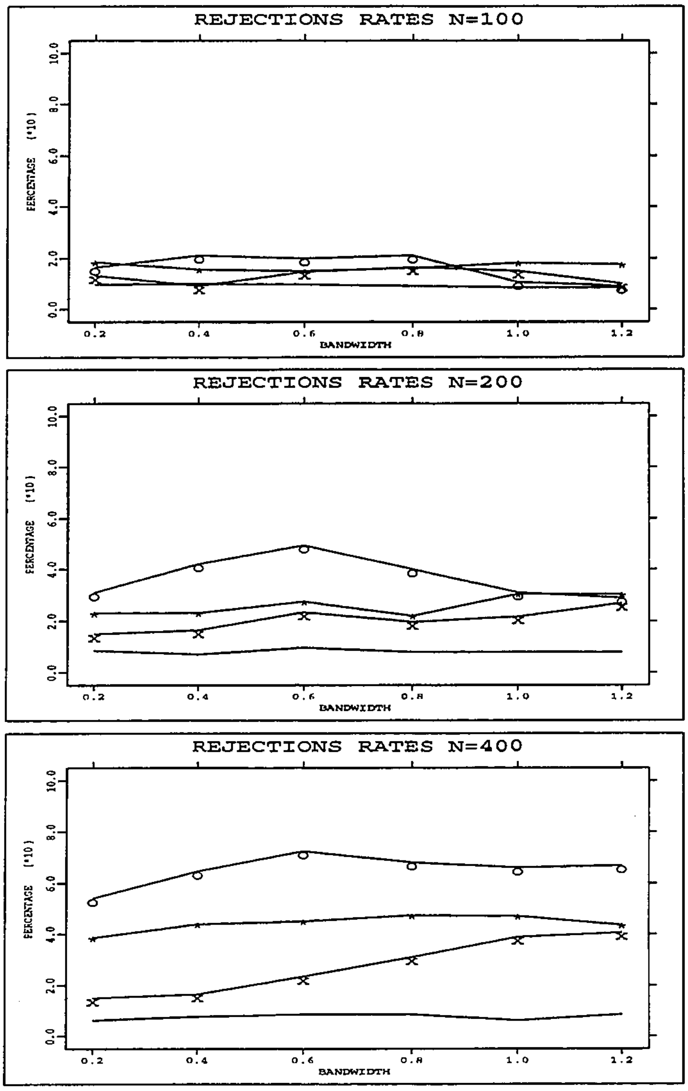

Figure 6.4: *Rates of rejection of one-sided MHH-test under the null (line) and under misspecification considering a logit with bump (line with circles), logit with heteroscedasticity (line with stars) and complementary log-log (line with crosses), for a nominal level of* 10%.

The rejection rate in the one–sided test are given by the percentage of the 200 runs for which the HH-test statistic was bigger than the asymptotical critical value. At the nominal level of 10% and 5% the critical values are

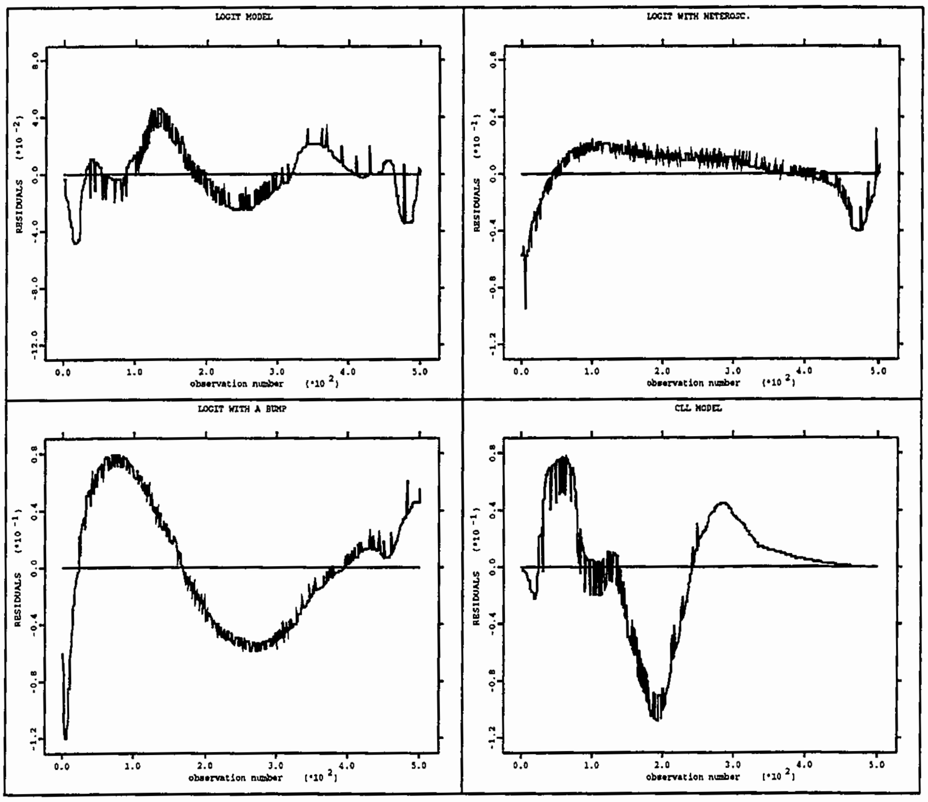

***Figure 6.5:** Residuals smoothed with the LOO kernel regression for* $h = 0.4$.

respectively 1.285 and 1.645.

Figure 6.3 shows the rejection rates of the statistic for the two–sided test at a 10% nominal level. For all sample sizes the empirical size (line) is better approximated than by the HH–test and seems to be almost the same for all bandwidths revealing some bandwith robustness of the test under the null. For samples with 100 observations the empirical power is very unsatisfactory, although it is much better than the empirical power obtained with the HH–test for the same situations (except for logit with heteroscedasticity when $h = 1.2$ and complementary log–log with $h = 0.2$). For the logit with bump (line with circles) the highest rejection rate is 16.5% as shows table 6.2 while with the HH–test the best rate for the same number of observations and the same model is 4.6%. The logit with heteroscedasticity (line with stars) has a best rejection rate 13.5% and the complementary log–log (line with crosses) 12.5%.

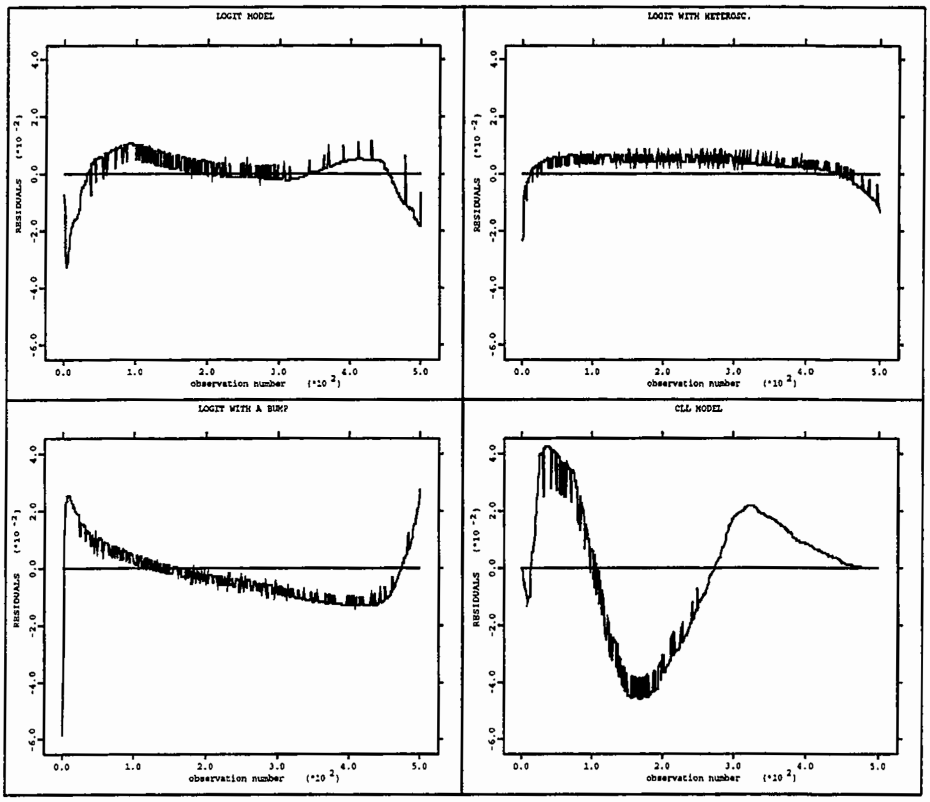

Figure 6.6: Residuals smoothed with the LOO kernel regression for $h = 1$.

For 200 observations the improvement in the power relatively to the HH–test is also very clear (except for the logit with heteroscedasticity when $h = 1.2$). The logit with bump attains the rejection rate of 38.5% while the best rate in the HH–test was 10.8%, the logit with heteroscedasticity achieves 26.5% of rejections and the complementary log–log 22% while with the HH–test the best rate was 8%.

In samples with 400 observations the best improvement in the empirical power is verified when data is generated from the complementary log–log. In this case the best rate of rejection is 36.5% while with the HH–test was 9.2%. For the logit with bump the best rate is 65.5% against 50.4% and for the logit with heteroscedasticity is 43.0% against 67.6%.

The one–sided test shows a better performance than the two–sided test.

	h=0.2	h=0.4	h=0.6	h=0.8	h=1.0	h=1.2
			n=100			
5th percent.	-1.37	-1.28	-1.16	-1.13	-1.14	-1.11
median	-0.30	-0.26	-0.30	-0.33	-0.30	-0.29
95th percent.	1.74	1.91	1.89	1.96	1.90	1.67
mean	-0.09	-0.07	-0.06	-0.07	-0.09	-0.12
st. deviation	1.02	0.96	0.95	0.95	0.93	0.89
			n=200			
5th percent.	-1.38	-1.24	-1.13	-1.10	-1.06	-1.04
median	-0.21	-0.26	-0.24	-0.34	-0.34	-0.37
95th percent.	1.55	1.51	1.75	1.73	1.96	1.85
mean	-0.11	-0.10	-0.09	-0.10	-0.11	-0.12
st. deviation	0.92	0.87	0.87	0.89	0.91	0.92
			n=400			
5th percent.	-1.45	-1.36	-1.16	-1.12	-1.13	-1.00
median	-0.28	-0.36	-0.18	-0.36	-0.41	-0.31
95th percent.	1.42	1.53	2.00	1.98	1.56	1.74
mean	-0.15	-0.19	-0.01	-0.12	-0.15	-0.05
st. deviation	0.98	0.94	0.93	0.93	0.92	0.89

Table 6.1: Percentiles, mean and standard deviation of the MHH–test statistic under correct specification

The empirical size is always very close to the nominal level of the test and the empirical power is greater in all situations, although the one–sided MHH–test cannot beat the two–sided HH–test with $h = 1.2$ for the logit with heteroscedasticity. However, mind that the rejections in that case were artificial in the sense they were due to a severe oversmoothing of the data.

Figures 6.5 and 6.6 show the residuals of the parametric regression (assuming a logit) smoothed by the LOO kernel regression estimator for bandwidth respectively of $h = 0.4$ and $h = 1$. The data are the same as in figure 4.15.

The smoothed residuals under correct mispecification (upper left), that is when data are generated by the logit model, have a very irregular behavior close to zero. Under misspecification it would be expected that residuals deviate more from zero. However for the logit with heteroscedasticity the residuals are still close to zero (elucidating the not so good performance of the test in this

	two–sided 10%	two–sided 5%	one–sided 10%	one–sided 5%	two–sided 10%	two–sided 5%	one–sided 10%	one–sided 5%
	Logit				Logit with a bump			
h = 0.2	6.5	3.5	9.5	5.5	12.5	9.0	16.5	12.0
	(1.74)	(1.30)	(2.07)	(1.61)	(2.34)	(2.02)	(2.62)	(2.30)
h = 0.4	8.5	4.5	10.0	8.5	16.5	11.5	21.0	16.5
	(1.97)	(1.47)	(2.12)	(1.97)	(2.62)	(2.26)	(2.88)	(2.62)
h = 0.6	8.0	4.0	9.5	8.0	14.5	14.0	20.0	14.5
	(1.92)	(1.39)	(2.07)	(1.92)	(2.49)	(2.45)	(2.83)	(2.49)
h = 0.8	6.0	4.5	9.0	6.0	16.0	11.0	21.0	16.0
	(1.68)	(1.47)	(2.02)	(1.68)	(2.59)	(2.21)	(2.88)	(2.59)
h = 1.0	6.5	3.5	8.5	6.5	6.5	5.5	10.5	6.5
	(1.74)	(1.30)	(1.97)	(1.74)	(1.74)	(1.61)	(2.17)	(1.74)
h = 1.2	5.5	3.0	8.5	5.5	5.0	4.0	9.0	5.0
	(1.61)	(1.21)	(1.97)	(1.61)	(1.54)	(1.39)	(2.02)	(1.54)
	Logit with Heteros.				CLL			
h = 0.2	11.0	7.5	18.5	10.0	13.0	9.0	11.0	7.5
	(2.21)	(1.86)	(2.75)	(2.21)	(2.38)	(2.02)	(2.21)	(1.86)
h = 0.4	11.0	8.5	15.5	10.5	9.0	7.0	8.0	4.5
	(2.21)	(1.97)	(2.56)	(2.17)	(2.02)	(1.80)	(1.92)	(1.47)
h = 0.6	12.0	9.0	15.0	12.0	14.5	10.0	10.5	8.0
	(2.30)	(2.02)	(2.52)	(2.30)	(2.49)	(2.21)	(2.17)	(1.92)
h = 0.8	12.5	11.0	16.0	12.5	16.5	12.5	12.5	9.5
	(2.34)	(2.21)	(2.59)	(2.34)	(2.62)	(2.34)	(2.34)	(2.07)
h = 1.0	11.5	9.0	18.0	11.5	15.0	11.0	11.0	7.5
	(2.26)	(2.02)	(2.72)	(2.26)	(2.52)	(2.21)	(2.21)	(1.86)
h = 1.2	13.5	10.0	17.5	13.5	10.0	7.0	7.0	4.5
	(2.42)	(2.21)	(2.69)	(2.42)	(2.21)	(1.80)	(1.80)	(1.47)

Table 6.2: Percentages of rejections of the MHH–test for 100 *observations. Standard errors between brackets.*

case) but with a tendency to be positive revealing some misspecification. The existence of misspecification is clear in the behavior of the smoothed residuals for the logit with bump and complementary log–log. Note that oversmoothing may hide the structure on the residuals and consequently may hide misspecification. That is the case for instance for the logit with a bump when $h = 1$ explaining why the test had a better performance for this model with $h = 0.6$.

	two–sided		one–sided		two–sided		one–sided	
	10%	5%	10%	5%	10%	5%	10%	5%
	Logit				Logit with a bump			
h = 0.2	6.5	2.5	8.5	4.5	24.0	20.0	31.0	24.0
	(1.74)	(1.10)	(1.97)	(1.47)	(3.02)	(2.83)	(3.27)	(3.02)
h = 0.4	4.0	3.0	7.0	4.0	35.0	30.5	42.0	35.0
	(1.39)	(1.21)	(1.80)	(1.39)	(3.37)	(3.26)	(3.49)	(3.37)
h = 0.6	6.0	2.5	9.5	6.0	38.5	34.5	49.5	38.5
	(1.68)	(1.10)	(2.07)	(1.68)	(3.44)	(3.36)	(3.54)	(3.44)
h = 0.8	5.5	4.5	8.0	5.5	32.0	25.5	40.0	32.0
	(1.61)	(1.47)	(1.92)	(1.61)	(3.30)	(3.08)	(3.46)	(3.30)
h = 1.0	5.0	4.5	8.0	5.0	25.5	18.5	31.0	25.5
	(1.54)	(1.47)	(1.92)	(1.54)	(3.08)	(2.75)	(3.27)	(3.08)
h = 1.2	5.5	4.0	8.0	5.5	21.0	14.5	29.0	21.0
	(1.61)	(1.39)	(1.92)	(1.61)	(2.88)	(2.49)	(3.21)	(2.88)
	Logit with Heteros.				CLL			
h = 0.2	16.0	12.5	23.0	16.0	11.0	6.0	15.0	9.5
	(2.59)	(2.34)	(2.98)	(2.59)	(2.21)	(1.68)	(2.52)	(2.07)
h = 0.4	16.5	13.0	23.0	16.5	9.5	7.5	16.5	9.5
	(2.62)	(2.38)	(2.98)	(2.62)	(2.07)	(1.86)	(2.62)	(2.07)
h = 0.6	20.0	16.5	27.5	20.0	14.5	11.0	23.5	14.5
	(2.83)	(2.62)	(3.16)	(2.83)	(2.49)	(2.21)	(3.00)	(2.49)
h = 0.8	20.0	17.0	22.0	20.0	15.0	12.0	19.5	15.0
	(2.83)	(2.66)	(2.93)	(2.83)	(2.52)	(2.30)	(2.80)	(2.52)
h = 1.0	26.0	23.0	30.5	26.0	17.5	15.0	21.5	17.5
	(3.10)	(2.98)	(3.26)	(3.10)	(2.69)	(2.52)	(2.90)	(2.69)
h = 1.2	26.5	21.0	30.5	26.5	22.0	19.5	27.0	22.0
	(3.12)	(2.88)	(3.26)	(3.12)	(2.93)	(2.80)	(3.14)	(2.93)

Table 6.3: Percentages of rejections of the MHH–test for 200 observations. Standard errors between brackets.

6.5 Comparison between the MHH–test and the BHH–test

This section compares the performance of the MHH–test relatively to the bootstrapped HH–test, the so–called BHH–test, introduced in chapter 5. With that

	two-sided 10%	two-sided 5%	one-sided 10%	one-sided 5%	two-sided 10%	two-sided 5%	one-sided 10%	one-sided 5%
	Logit				Logit with a bump			
h = 0.2	5.0	2.5	6.0	3.5	48.0	43.0	54.0	47.5
	(1.54)	(1.10)	(1.68)	(1.30)	(3.53)	(3.50)	(3.52)	(3.53)
h = 0.4	4.5	2.0	7.5	3.5	57.5	53.5	64.5	57.5
	(1.47)	(0.99)	(1.86)	(1.30)	(3.50)	(3.53)	(3.38)	(3.50)
h = 0.6	6.5	5.0	8.5	6.5	65.5	58.5	72.5	65.5
	(1.74)	(1.54)	(1.97)	(1.74)	(3.36)	(3.48)	(3.16)	(3.36)
h = 0.8	6.5	5.0	8.5	6.5	59.0	55.0	68.0	59.0
	(1.74)	(1.54)	(1.97)	(1.74)	(3.48)	(3.52)	(3.30)	(3.48)
h = 1.0	4.5	3.0	6.0	4.5	61.5	54.5	66.0	61.5
	(1.47)	(1.21)	(1.68)	(1.47)	(3.44)	(3.52)	(3.35)	(3.44)
h = 1.2	5.5	3.5	8.5	5.5	58.5	52.5	67.0	58.5
	(1.61)	(1.30)	(1.97)	(1.61)	(3.48)	(3.53)	(3.32)	(3.48)
	Logit with Heteros.				CLL			
h = 0.2	31.0	24.5	38.5	31.0	11.0	6.0	15.0	9.5
	(3.27)	(3.04)	(3.44)	(3.27)	(2.21)	(1.68)	(2.52)	(2.07)
h = 0.4	34.5	27.0	44.0	34.5	9.5	7.5	16.5	9.5
	(3.36)	(3.14)	(3.51)	(3.36)	(2.07)	(1.86)	(2.62)	(2.07)
h = 0.6	38.0	33.5	45.0	38.0	14.5	11.0	23.5	14.5
	(3.43)	(3.34)	(3.52)	(3.43)	(2.49)	(2.21)	(3.00)	(2.49)
h = 0.8	43.0	38.5	47.5	43.0	24.0	21.0	31.0	24.0
	(3.50)	(3.44)	(3.53)	(3.50)	(3.02)	(2.88)	(3.27)	(3.02)
h = 1.0	39.0	35.0	47.0	39.0	33.5	23.0	39.0	33.5
	(3.45)	(3.37)	(3.53)	(3.45)	(3.34)	(2.98)	(3.45)	(3.34)
h = 1.2	37.5	34.5	43.5	37.5	36.5	30.5	40.5	36.5
	(3.42)	(3.36)	(3.50)	(3.42)	(3.40)	(3.26)	(3.47)	(3.40)

Table 6.4: Percentages of rejections of the MHH–test for 400 *observations. Standard errors between brackets.*

purpose a simulation study was performed for 200 pseudo samples. In each sample the BHH–test was calculated according to algorithm 5.1 of chapter 5 considering bandwiths $h = 0.4$ and $h = 1$. Empirical size and empirical power of the BHH–test were determined according to algorithm 5.2 of chapter 5.

The results obtained for samples with 100 observations are shown in ta-

	two–sided		one–sided	
	BHH	MHH	BHH	MHH
		Logit		
h = 0.4	8.0	8.5	9.5	10.0
	(1.92)	(1.97)	(2.07)	(2.12)
h = 1.0	6.5	6.5	11.0	8.5
	(1.74)	(1.74)	(2.21)	(1.97)
		Logit with a bump		
h = 0.4	20.5	16.5	26.0	21.0
	(2.85)	(2.62)	(3.10)	(2.88)
h = 1.0	29.5	6.5	39.0	10.5
	(3.22)	(1.74)	(3.45)	(2.17)
		Logit with Heteros.		
h = 0.4	10.0	11.0	14.5	15.5
	(2.12)	(2.21)	(2.49)	(2.56)
h = 1.0	8.5	11.5	8.0	18.0
	(1.97)	(2.26)	(1.92)	(2.72)
		CLL		
h = 0.4	14.0	9.0	10.0	8.0
	(2.45)	(2.02)	(2.12)	(1.92)
h = 1.0	16.5	15.0	18.5	11.0
	(2.62)	(2.52)	(2.75)	(2.21)

Table 6.5: *Percentages of rejections of the BHH–test and MHH–test for* 100 *observations and nominal size of* 10%*. Standard errors between brackets.*

ble 6.5 while for samples with 200 observations can be seen in table 6.6. Samples of greater size were not experimented given the heavy calculations required by the bootstrap.

There is not a clear superiority of one of the tests procedures. When the data is generated by the logit with heteroscedasticidy the MHH–test proved to perform better regardless the sample size or the bandwidth. For $h = 0.4$ and samples with 100 observations the performance achieved by both tests is close with a slight superiority of the BHH–test for the Logit with a bump and the CLL models which may not compensate the greater expense in calculations of this test. When the sample size is 200 and $h = 0.4$ the MHH–test presents a better performance. For bandwidth $h = 1$ the BHH–test shows better results in

	two–sided		one–sided	
	BHH	MHH	BHH	MHH
	Logit			
h = 0.4	10.0	4.0	7.5	7.0
	(2.12)	(1.39)	(1.86)	(1.80)
h = 1.0	9.5	5.0	16.0	8.0
	(2.07)	(1.54)	(2.59)	(1.92)
	Logit with a bump			
h = 0.4	27.0	35.0	39.5	42.0
	(3.14)	(3.37)	(3.46)	(3.49)
h = 1.0	44.0	25.5	55.0	31.0
	(3.51)	(3.08)	(3.52)	(3.27)
	Logit with Heteros.			
h = 0.4	12.5	16.5	21.0	23.0
	(2.34)	(2.62)	(2.88)	(2.98)
h = 1.0	13.5	26.0	2.5	30.5
	(2.42)	(3.10)	(1.10)	(3.26)
	CLL			
h = 0.4	12.0	9.5	13.5	16.5
	(2.30)	(2.07)	(2.42)	(2.62)
h = 1.0	20.5	17.5	33.5	21.5
	(2.85)	(2.69)	(3.34)	(2.90)

Table 6.6: *Percentages of rejections of the BHH–test and MHH–test for* 200 *observations and nominal size of* 10%.

empirical power when data are generated from a logit with a bump and a CLL. However, for samples with 200 observations the one-sided BHH–test shows a too great empirical size.

The results do not allow to elect a test procedure as the preferable to use in all situations. One possible option consists in using the MHH–test if the computing environment allows to make calculations with matrices with many rows and columns as observations in the data set. In this situation the MHH–test is preferable because the calculations should be faster and more direct (e.g. can be made in XploRe 3 with procedure `MODHHTST`) and should not perform worst than the BHH–test. If possible both tests should be claculated. When the sample size disallow the use of the MHH–test then usually the BHH–test can be still implemented because though it needs more calculations it consumes less RAM because it works with matrices of smaller size.

6.6 Concluding Remarks

The results described before allow to draw the following conclusions. The MHH–test approximates considerably better the empirical size to the nominal size of the test than the HH–test. Moreover, the size is pratically independent of the bandwidth and is not expected to explode provoking a too high rate of false rejections (Proença and Ritter (1994) performed simulations for samples with 1000 observations revealing also a stable size of the test).

The MHH–test improves in general the empirical power relatively to the HH–test specially for the complementary log–log. However it is not able to beat the performance of the two–sided HH–test with large bandwidth for the logit with heteroscedasticity. This may be irrelevant considering that the good performance there was artificial because it was due essentially to a significant oversmoothing of the data. In practice oversmoothing can be dangerous giving that it can lead to the rejection of a well specified model and consequently should be avoided.

The superior behavior of the MHH–test relatively to the HH–test is due to the reduction of the negative bias and the bandwidth dependency. To obtain these features is vital to use the bias correction in (6.6) and the corrected variance defined by (6.7). Some simulations performed within the work of this thesis have shown that the MHH–test without the bias correction does not perform much better than the HH–test due to the presence of the negative bias. To not burden the text these results were not included here.

The one–sided test is specially advised to use given that has a better performance than the two–sided test in all situations analysed.

The power of the test depends on the bandwidth. This fact should be understanded as a feature. An examination of the behavior of the test statistic versus the bandwidth used in the smoothing operation which indicates rejection of the parametric assumed model in some part and acceptance in another is by no means contradictory. On the contrary, it provides information about the way the alternative differs from the null hypothesis.

The MHH–test proved in general to perform at least as good as the bootstrap correction of the HH–test, the BHH–test. However the MHH–test involves calculations with squared matrices with the same dimension as the sample size. Therefore, its calculation needs great amounts of RAM as the sample size grows. For example, for samples with 400 observations it needs almost 16 of RAM. This means that for samples with thousands of observations the MHH–test is difficult to apply in practice. In this case the practioneer should use the BHH–test to avoid the pernicious effect of the negative bias in the HH–statistic and improve the power of the test. Although the BHH–test incurs in more calculations than the MHH–test it needs smaller amounts of RAM.

The dimension of the two data sets under study in this dissertation, the travel-mode choice and credit-scoring, made impossible the calculation of the MHH–test with the computational means available during this work. However, one expects that the possible conclusions the MHH–test would reach would not be too different from those obtained with the BHH–test in chapter 5. To see an illustration of how the MHH–test behaves with real data proceed to next chapter namely to the analysis of the data about unemployment after apprenticeship.

Chapter 7

Empirical Applications

7.1 Introduction

This chapter has the aim to exemplify how the techniques introduced in this dissertation can be applied in analyzing a real data set with binary responses.

Two data sets were used along all the chapters of this work to illustrate the methods in study. These data sets were already used and analyzed, though in different contexts, by respectively Horowitz (1993) and Fahrmeir and Tutz (1994). The conclusions reached by the methods introduced in this thesis are in conformity with the results obtained by those authors.

This chapter analyzes two different data sets. One, concerns the problem of unemployment after apprenticeship. These data were analyzed in Proença and Werwatz (1994) and was elaborated during the research performed for this thesis. The study presented here follows closely the mentioned work. Details about the procedures used to make the calculations in XploRe are omitted here. The reader who is interested in those details should see Proença and Werwatz (1994).

The other data set consists in another example of a credit-scoring problem and it is a subsample of a an original data set analyzed at Institut de Statistique, Université Catholique de Louvain, by Cécile Denis and Agnés Chalon under the supervision of Professor Jean-Marie Rolin. They have used a Bayesian approach with different proposes than those guiding this work.

The aim here is to test the adequacy of the logit model assuming that the linear index is correctly specified. That is, to compare the logit model against the semiparametric S.I.M. in (2.1). After analyzing the parametric and semiparametric estimates for the coefficients of the index function uniform confidence bands will be calculated together with the HH–test and its corrections BHH–test and MHH–test.

7.2 Modelling Unemployment after Apprenticeship

This section is dedicated to the analysis of the data set about unemployment after apprenticeship. The data was kindly provided by Axel Werwartz. A description of the data set is presented below.

7.2.1 The Data

This section gives a brief description of the data used in this study. All data used in this project refer to the former West-Germany.

For the purposes of this study a sample of 462 individuals was extracted from the first nine waves (1984-1992) of the GSOEP (German socioeconomic panel). For a detailed description of the GSOEP see Projektgruppe (1991). Each year, respondents are asked whether they have completed an apprenticeship in the previous year. Those who answered "yes" to this question sometime between 1985 and 1992 were included in our sample.
The dependent variable **UNEMP** takes on the value "1" if an individual is registered as unemployed at the time of the survey in the year following the completion of the apprenticeship. It takes on the value "0" if the individual is employed. Proença and Werwatz (1994) assume that the probability than **UNEMP** takes on the value "1" is related to the following set of explanatory variables, summarized in table 7.1:

SCHOOLING and **AGE** are trying to measure general human capital and are expected to have a negative effect on the probability of being unemployed. **EARNINGS** is supposed to capture "the value of an apprenticeship". Again, one expects a negative coefficient for this variable. The effect of **CITYSIZE** is not clear cut, a priori, but one can make a case that, ceteris paribus, larger cities offer more employment opportunities and therefore a negative coefficient should be expected. It has been observed that for small firms (especially in the artisan sector) the number of apprenticeship positions provided exceeds the number of workers retained after the apprenticeship is completed. Hence,

Variable	Definition/Comments
AGE	age of the respondent in the year the apprenticeship was completed
SCHOOLING	years of schooling
EARNINGS	gross monthly earnings as an apprentice
CITYSIZE	size of the city the respondent lives in at the time the apprenticeship was completed
FIRMSIZE	size of the firm where the respondent was apprenticed
DEGREE	percentage of people apprenticed in the occupation of the respondent divided by the percentage of people employed in this occupation in the entire economy
URATE	unemployment rate in the state the respondent lived in during the year the apprenticeship was completed

Table 7.1: Explanatory Variables for the unemployment after apprenticeship data.

FIRMSIZE is likely to have a negative effect on the probability of being unemployed. **DEGREE** and **URATE** are variables that are generated from information provided by Germany's federal statistical bureau, the *Statistisches Bundesamt*. **URATE** is supposed to capture the overall employment situation in the state the respondent lived in at the time he or she completed an apprenticeship. Clearly, one expects a positive coefficient for this variable. **DEGREE** derives its name from the fact that by definition the variable measures the degree to which an occupation is "overapprenticed". That is, suppose that 20% of all apprentices are trained as mechanics but only 10% of all workers in the entire German economy work as mechanics. Then **DEGREE** will take on the value $\frac{0.2}{0.1} = 2$. Hence, if a respondent has completed an apprenticeship that is "overapprenticed" in the sense just described then **DEGREE** will take on values greater than "1". A significant positive coefficient of this variable would indicate that those who completed an apprenticeship in overapprenticed occupations face a higher probability of becoming unemployed after the apprenticeship is finished.

A final remark to stress that this data is not a panel data because each year different individuals were observed.

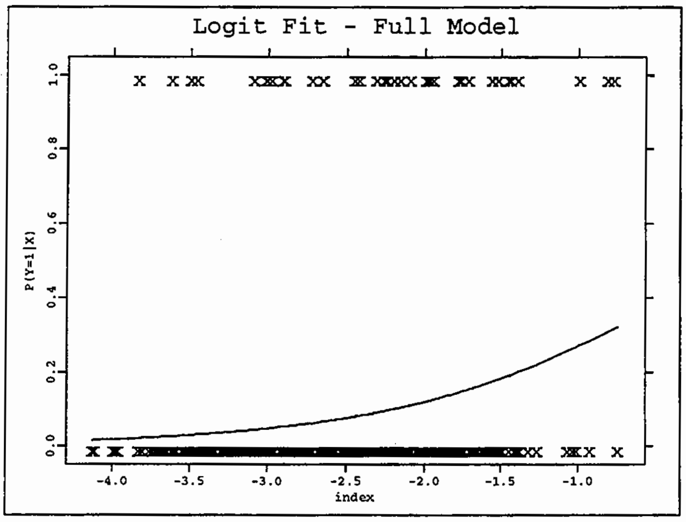

Figure 7.1: The logit fit for the full model. The data points are identified by crosses. Unemployment data.

7.2.2 The Parametric Fit

To start with, a logit model was estimated including all explanatory variables described above plus an intercept term. This model will be referred to as the "full model". The results of the parametric fit were obtained using XploRe 3. They are shown in table 7.2. The plot of the logit regression is shown in figure 7.1.

The coefficients of all variables, with the exception of **AGE** and **FIRMSIZE**, have the expected signs. The coefficient of **DEGREE** is negative but highly insignificant. In fact, most coefficients are statistically insignificant at the 5% level. Hence, several variables should be dropped on the basis of their low t-ratios.

Note that this strategy is open to criticism. First of all, McCullagh and Nelder (1989) have pointed out that in the binary logit model the true distribution of the t-statistic may be poorly approximated by the t–distribution for samples of the size encountered in this application. Moreover, t-tests hinge on the assumption that the logit link is correctly specified. This assumption has not been tested, yet.

Still, **AGE**, **SCHOOLING**, and **FIRMSIZE** were eliminated from the

	coeff.	st. errors
INTERCEPT	-4.27032	2.25014
AGE	0.03718	0.11426
SCHOOLING	-0.01280	0.17347
EARNINGS	-0.00070	0.00098
CITYSIZE	-0.00048	0.00042
FIRMSIZE	0.00025	0.00265
DEGREE	-0.00116	0.00200
URATE	0.23398	0.06504
454	Degrees of freedom	
237.394	Deviance	
490.030	Pearson's χ^2	

Table 7.2: Results of the Logit Fit – Full Model. Unemployment data.

	coeff.	st. errors
INTERCEPT	-3.65076	0.99971
EARNINGS	-0.00065	0.00094
CITYSIZE	-0.00046	0.00041
DEGREE	-0.00116	0.00199
URATE	0.23540	0.06503
457	Degrees of freedom	
237.511	Deviance	
488.129	Pearson's χ^2	

Table 7.3: Results of the logit fit – restricted model. Unemployment data.

model. **EARNINGS** and **DEGREE** were kept in the model despite their low t-values because these variables are particularly interesting from an economic point of view. This leads to the "restricted model" with **DEGREE**, **EARNINGS**, **CITYSIZE** and **URATE** as explanatory variables. The results of the logit fit for the restricted model are given in table 7.3. The respective plot has shown to be very similar to the plot of the logit fit for the full model as can be seen in for instance in Figure 7.3.

As in the full model, **URATE** has a significant positive effect and appears to be the most important predictor of the probability of being unemployed.

In the following section, estimates for the semiparametric fit of the restricted

	Logit	WADE			Klein-Spady
		h=1.25	h=1.5	h=1.75	
EARNINGS	-1.00	-1.00	-1.00	-1.00	-1.00
	(1.45)	(1.18)	(1.43)	(1.64)	(0.60)
CITYSIZE	-0.72	-0.47	-0.66	-0.81	-0.64
	(0.63)	(0.27)	(0.37)	(0.45)	(0.09)
DEGREE	-1.79	-1.52	-1.93	-2.25	-2.16
	(3.06)	(2.63)	(3.22)	(3.73)	(1.20)
URATE	363.03	245.74	319.47	384.45	314.60
	(100.05)	(85.65)	(108.33)	(127.94)	(67.81)

Table 7.4: Results of the Semiparametric fit – restricted model. Standard errors between brackets. Unemployment data.

model will be presented and compared with the logit estimates obtained above.

7.2.3 The Semiparametric Fit

The semiparametric fit was estimated by the methods of WADE and Klein and Spady (1993) introduced in chapter 2. The WADE estimates were obtained with XploRe 3 while the others were calculated using GAUSS as in chapter 2.

For technical reasons, a Mahalanobis transformation has to be applied to the explanatory variables in order to eliminate correlation and to standardize. After estimating the coefficients of the transformed variables by WADE one can get coefficient estimates for the untransformed variables by postmultiplying the vector of coefficients of the transformed variables by the transformation matrix. Details about these calculations can be seen in Proença and Werwatz (1994).

The coefficient estimates of the untransformed variables are reported in table 7.4. The respective standard errors are inside brackets.

Recall that in the semiparametric model of the form (2.1) the intercept of the index $x^T\beta$ has to be subsumed into the definition of the link function as it was explained in chapter 2. Hence, no estimate for β_0 will be obtained from the semiparametric fit. Moreover, the scale of the coefficients was normalized by dividing all coefficients by the absolute value of the coefficient of **EARNINGS**. The same scale normalization was applied to the logit estimates of table 7.3. The normalized logit estimates are reported for purposes of comparison.

The signs of the estimated coefficients of all explanatory variables are neither varying with the bandwidth chosen in the second step of the semiparametric estimation nor do they differ between the logit model and the single index model. Note that for $h = 1.5$ the coefficient estimates of the two models are quite close. Hence, it appears that the logit link is not grossly misspecified. The following sections take a closer look at this issue.

7.2.4 Testing the Adequacy of the Logit Link

In this section, the adequacy of the link function of the logit model is going to be tested using the tools under study in this dissertation. More specifically, these tools consist of uniform confidence bands and the specification test developed by Horowitz and Härdle, the HH–test, together with its corrections introduced in this thesis which lead to the BHH–test and the MHH–test studied respectively in chapter 5 and chapter 6.

To construct the uniform confidence band the index $x^T\beta$ is estimated semiparametrically (see table 7.4 for the coefficient estimates). With β estimated semiparametrically one believes that the confidence limits are more robust in case of misspecification. However, comparing graphically the confidence limits with the parametric fit is more complicated in this situation because the semiparametric link function is not defined on the same scale as the Logit link. To overcome this problem algorithm 2.1 presented in chapter 2.

The uniform confidence band was obtained with XploRe 3 running procedure `UNIFBAND`. Figure 7.2 shows the resulting plot.

The confidence band in Figure 7.2 unequivocally speak in favor of the logit link. Yet, as it was focused in chapter 2 this procedure is rather an informal way to test the adequacy of $F(u) = 1/[1 + exp(-u)]$.

The HH–test was applied to the restricted model for a grid of bandwidths ranging from 0.5 to 1.75. The value of the statistic for each bandwidth along with the respective p–values are shown in table 7.5. For all bandwidths tried the test does not reject the logit link. Figure 7.3 shows the Logit fit along with the leave–one–out (LOO) estimate used in the HH–test for a bandwidth of $h = 1$. The corresponding plot with the classic Nadaraya–Watson estimate in place of the LOO estimate looks very similar and is not shown here. The parametric and the semiparametric fit are very close, suggesting a correct specification of the logit link.

The MHH–test was also calculated for the restricted model. Because this test is computationally expensive it was calculated only for three bandwidth

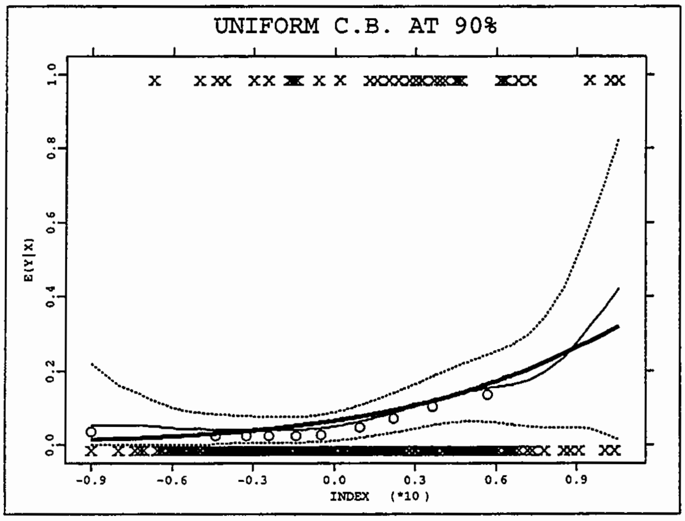

Figure 7.2: The Logit fit (thick line), the semiparametric estimate (line with circles) and the uniform confidence bands (broken line). The index was estimated by WADE. Unemployment data.

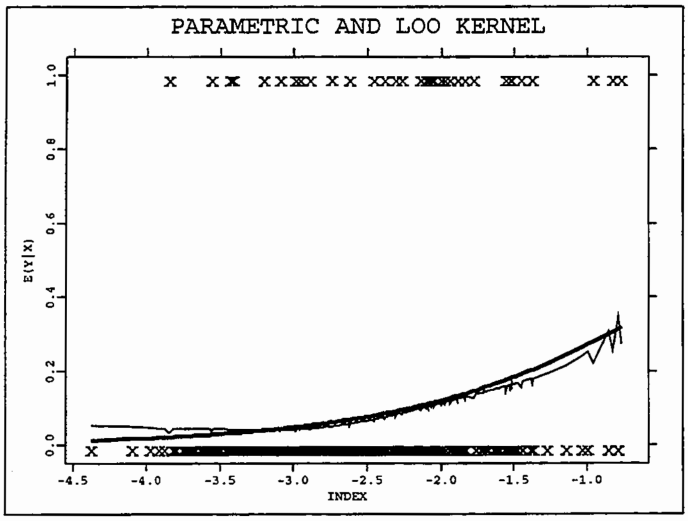

Figure 7.3: The Logit fit for the restricted model (thick line) and the LOO semiparametric fit used in the HH–test (line). The data points are identified by crosses. Unemployment data.

h	HH-test	p–value
0.50	-1.0741	0.141
0.75	-1.0210	0.154
1.00	-0.9276	0.177
1.25	-0.8764	0.190
1.50	-0.8449	0.199
1.75	-0.7862	0.216
	Modified HH–test	
0.75	-0.6224	0.329
1.00	-0.4716	0.357
1.25	-0.4449	0.361

Table 7.5: Results for the HH–test and Modified HH–test. Unemployment data.

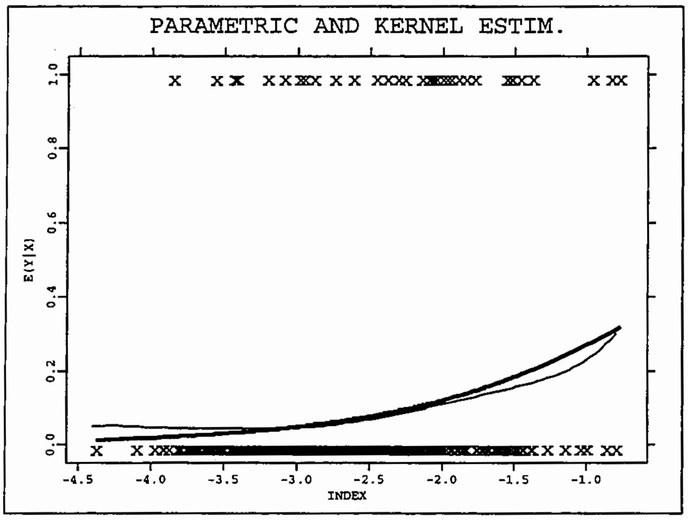

Figure 7.4: The Logit fit for the restricted model (thick line) and the semi-parametric kernel regression (line). The data points are identified by crosses. Unemployment data.

values. However, the test is bandwidth robust under the null and consequently it is likely that for other bandwidth values we would obtain very similar results.

Figure 7.5 shows the residuals of the parametric fit together with their smoothed counterparts to which the leave–one–out estimator was applied. The

h	HH-test	5% c.v.	90% c.v.	95% c.v.
0.3	-0.7567	-1.13	0.66	1.02
0.4	-0.6551	-1.08	0.49	0.91
0.5	-0.5730	-1.33	0.87	1.38

Table 7.6: *Results for the HH-test and bootstrap critical values. Unemployment data.*

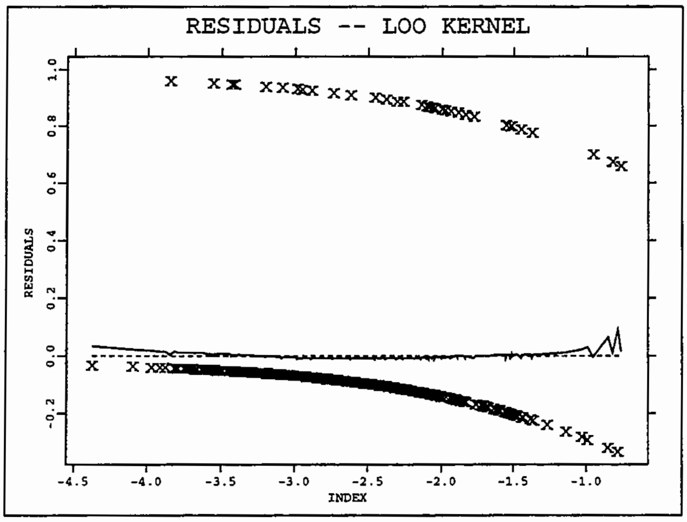

Figure 7.5: *The residuals of the Logit fit for the restricted model (crosses) and the residuals smoothed by LOO (line). Unemployment data.*

bandwidth was set to $h = 1$. The smoothed residuals are practically equal to zero indicating that the logit link is well specified. The corresponding picture with the residuals smoothed by the Nadaraya–Watson estimator is very similar and is not included for space reasons. The value of the test statistic indicates that the logit link should not be rejected. Note that the modified statistic has a greater value than the HH–test statistic reflecting the bias correction.

Table 7.6 shows the bootstrap corrected critical values for the HH–test together with the HH–statistic. Note that the bandwidth is different because this results were obtained using the Gaussian kernel while for the results presented before the smoothing was performed with the Quartic kernel. This problem was focused already in chapter 5. The values in the mentioned table indicate also that the logit fit should not be rejected.

7.3 The Credit–Scoring problem

This section is devoted to the analysis of the data set about credit-scoring referred to in the introduction of this chapter. The same structure of analysis used with the unemployment data is also pursued here except that given the dimension of the sample the MHH–test was not possible to be calculated.

A brief description of the data used in this section is given in the following.

7.3.1 The Data

The data set contains 1812 observations referring to customers leaving in the Flemish region of Belgium that make orders by phone to a given firm. The ordered products are sent to the customers and they pay the bill (if they decide to keep the product) by several installments.

The dependent variable **CLIENT** takes on the value "1" if a client is a good client, that is, pays the installments with no serious delays. It takes on the value "0" otherwise.

The explanatory variables that were considered to be related to the probability of a client to be good are summarized in Table 7.7. **TOTCRED**, the logarithm of the total amount of credit borrowed by the client measures the importance of the debt and it is expected that it contributes negatively to the probability of being a good client. **AGE** characterizes the client and one expects as older the client is there are greater chances of having a better economic status and to be more responsible. Therefore this variable should have a coefficient with positive sign. **PRICE** is the price of the most expensive commodity in the order and can be considered a proxy to measure the economic status of the client. Consequently, its coefficient should be positive. **MULTIOWNER** takes on the value "1" if more than one person has signed the same order and it is equal to "0" if only one person is responsible for the order. One should expect a negative contribute of this variable for the probability of being a good client because if more than one person are responsible for the same debt each one feels less obliged to fulfill its payment. **MULTIORDERS** takes on the value "1" if more than one order was done by the same client at the same day and takes on the value "0" otherwise. A negative value for the coefficient of this variable is anticipated because if a client makes more than one order at

Variable	Definition/Comments
TOTCRED	logarithm of total amount of credit borrowed
AGE	age of the client
PRICE	price of the most expensive good in the order
MULTIOWNER	more than one person owning the same order
MULTIORDERS	more than one order by the same client

Table 7.7: Explanatory Variables for the credit–scoring data.

the same day one expects that this client has a less responsible behavior and consequently feels less dutiful to fulfill the payments.

An additional remark concerning the use of the logarithm transformation in the total amount of credit. Some fits were obtained with the total amount of credit on the original levels but their results were not as good as those obtained when the logarithm transformation was in use.

7.3.2 The Parametric Fit

A logit model was adjusted to the data in study including all the explanatory variables in Table 7.7 plus an intercept term. The resulting fitted probability curve for a client to be a good one can be seen in Figure 7.6. One remarks that this probability is always over 0.6.

The estimated coefficients and respective estimated standard errors are included in Table 7.8. The signal of all estimates is in agree to the anticipations made before.

Analyzing the t–statistics for each coefficient one may conclude that the variables **PRICE** and **MULTIORDERS** are not important to explain the probability of being a good client. Therefore, another logistic regression was run in a data set without including those variables. The results can be seen in Table 7.9.

For the restricted model all coefficients are statistically different from zero based on the values of the t–statistics. Again, they have the expected signs.

Figure 7.7 shows the plot of the logit fit for the restricted model. There is not much difference between this fit and the logit fit for the model with all variables except that the lower bound for the probability is greater in the restricted model. The restricted fit is preferable attending at the parsimonious

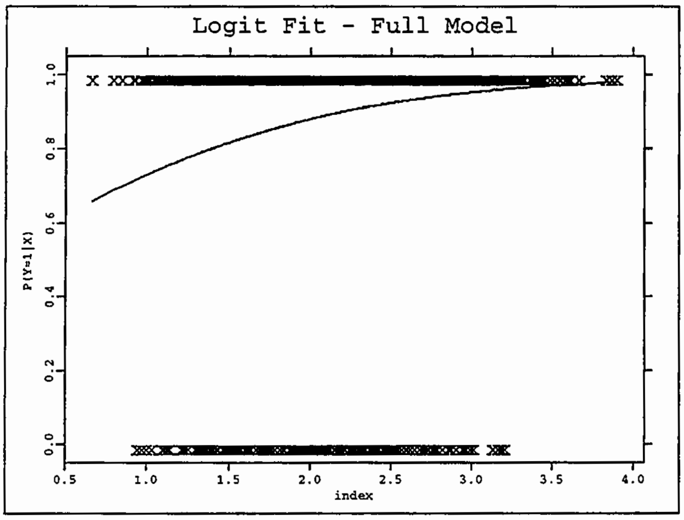

Figure 7.6: The logit fit for the full model. The data points are identified by crosses. Credit-scoring data.

	coeff.	st. errors
INTERCEPT	3.64554	1.36303
TOTCRED	-0.30001	0.16280
AGE	0.02744	0.00557
PRICE	0.00001	0.00002
MULTIOWNER	-0.46530	0.16408
MULTIORDERS	-0.41105	0.34746
1806	Degrees of freedom	
1178.4	Deviance	
1798.8	Pearson's χ^2	

Table 7.8: Results of the Logit Fit – Full Model. Credit-scoring data.

principle.

7.3.3 The Semiparametric Fit

Thinking in a possible misspecification of the logit link the coefficients for the explanatory variables **TOTCRED**, **AGE** and **MULTIOWNER** were

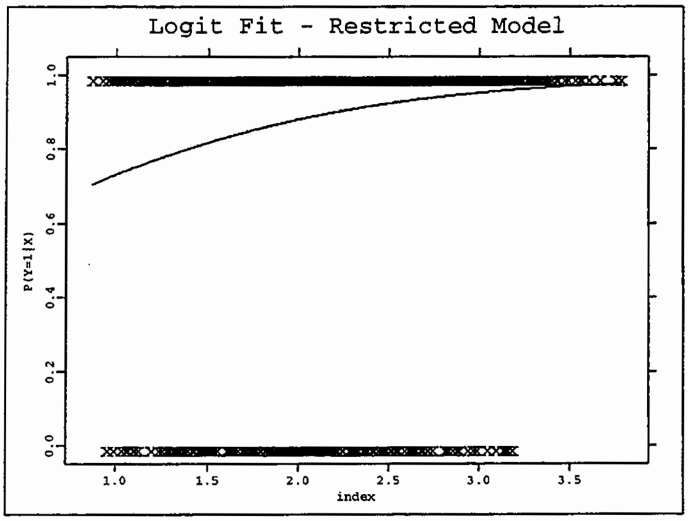

Figure 7.7: The logit fit for the restricted model. The data points are identified by crosses. Credit–scoring data.

	coeff.	st. errors
INTERCEPT	2.98927	0.89093
TOTCRED	-0.21892	0.09222
AGE	0.02753	0.00556
MULTIOWNER	-0.47042	0.16385
1808	Degrees of freedom	
1180.1	Deviance	
1808.6	Pearson's χ^2	

Table 7.9: Results of the logit fit – restricted model. Credit–scoring data

estimated semiparametrically by the method of Klein and Spady (1993). The results are in Table 7.10.

The semiparametric and parametric estimates have both the same signs. However, the standard error of the semiparametric estimate for the coefficient of **MULTIOWNER** shows that this variable is not statistically significant to explain the probability of being a good client.

Next, the parametric and semiparametric fits will be compared in order to

	Logit	st. errors	Klein-Spady	st. errors
TOTCRED	-1.00	0.42	-1.00	0.28
AGE	0.13	0.03	0.17	0.04
MULTIOWNER	-2.15	0.75	-1.51	1.45

Table 7.10: Results of the Semiparametric fit – restricted model. Credit-scoring data.

conclude if the logit model is adequate to fit the credit-scoring curve for these data.

7.3.4 Testing the Adequacy of the Logit Link

In order to assess the adequacy of the logit link uniform confidence bands and the HH–test with the bootstrap–corrected critical values are calculated.

The uniform confidence bands were calculated using each the semiparametrically and the parametrically estimated index. The respective plots can be seen in Figure 7.8 and Figure 7.9. Note that the amount of smoothing is different in both plots. In Figure 7.9 the amount of smoothing is greater. Note also that the bandwidth values used in each plot are not comparable because the respective estimated indexes are in a different scale.

In both plots the logit fit lies inside the confidence bands suggesting that it is well specified in this problem. However, one can notice that the left tail of the semiparametrically fitted probability curve lies clearly over from the parametric curve indicating that possibly the lower bound of being a good client is slightly greater than the one shown by the logit fit.

The HH–test is calculated for different bandwidths. Figure 7.10 shows the shape of the kernel regression for those bandwidths. Bandwidth $h = 0.08$ is moderately undersmoothing the data while bandwidth $h = 0.3$ is moderately oversmoothing it.

The results for the HH–statistic and bootstrap critical values are contained in Table 7.11. The 90% and 95% bootstrap critical values reflect the typical negative bias of the HH–test. The HH–statistic assumes negative values for all bandwidths. For $h = 0.08$ the test indicates rejection of the logit link whether one considers the standard normal critical values or bootstrap critical values. Curiously, for $h = 0.30$ the BHH–test leads to rejection while with

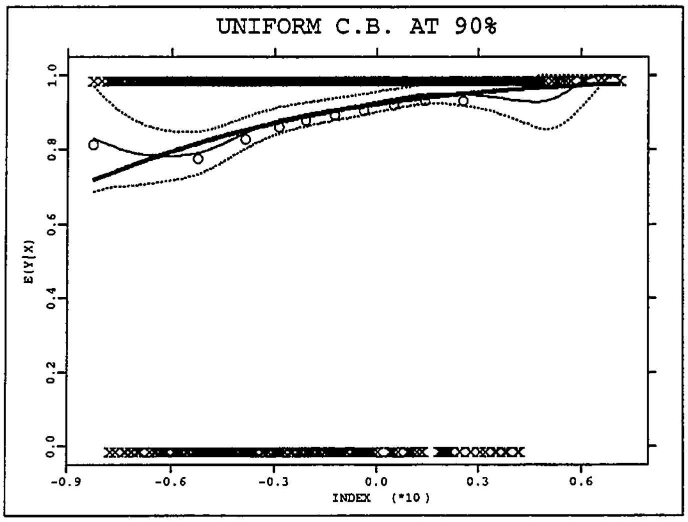

Figure 7.8: *The Logit fit (thick line), the semiparametric regression (h=2) (line with circles) and the uniform confidence bands (broken line). The index was estimated semiparametrically by maximum quasi-likelihood. Credit data.*

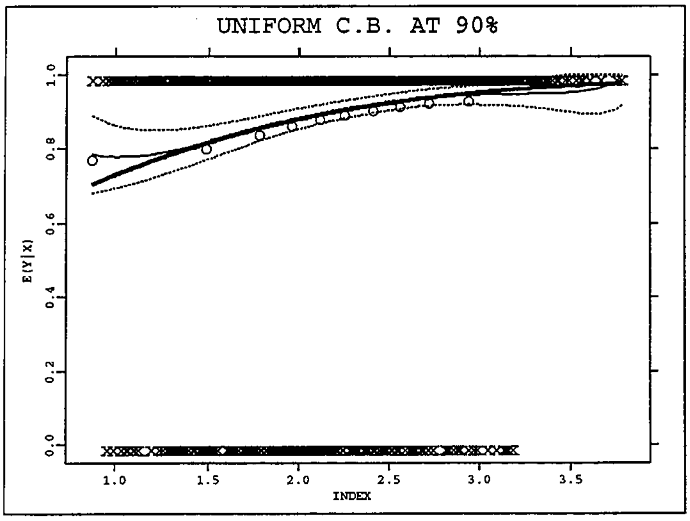

Figure 7.9: *The Logit fit (thick line), the semiparametric regression (h=0.3) (line with circles) and the uniform confidence bands (broken line). The index was estimated parametrically. Credit data.*

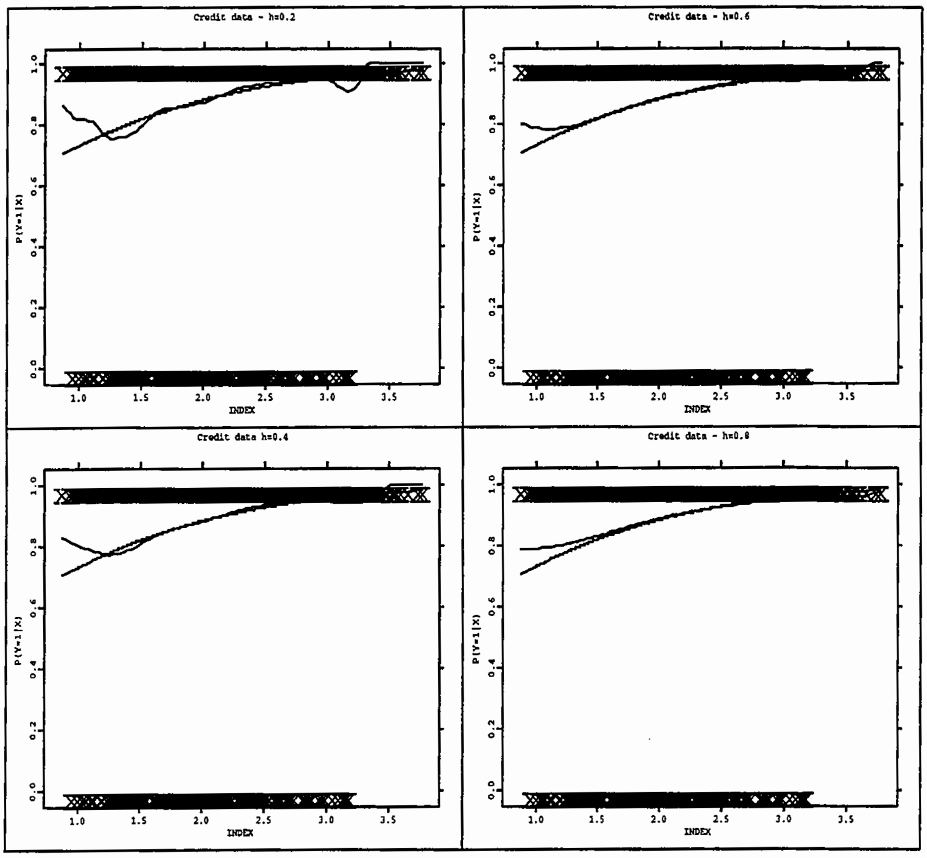

Figure 7.10: The Logit fit and the semiparametric kernel regression for the restricted model. The data points are identified by crosses. Upper left – h=0.08; lower left – h=0.15; upper right – h=0.23; lower right – h=0.30

the standard normal critical values the test indicates no rejection of the logit link. Is very much possible that those rejections are artificially created by too much undersmoothing and oversmoothing. Therefore, and considering also the result of the uniform confidence band, one should be more inclined to accept the correctness of the logit link in this problem.

h	HH–test	5% c.v.	90% c.v.	95% c.v.
0.08	-1.6539	-1.49	0.69	1.27
0.15	-1.3292	-1.38	0.34	0.93
0.23	-1.3113	-1.38	0.04	0.36
0.30	-1.4591	-1.43	0.05	0.50

Table 7.11: Results for the HH–test and bootstrap critical values. Credit-scoring data.

7.4 Concluding Remarks

In this section it was illustrated how the adequacy of the link function of a parametric binary choice model can be evaluated using the statistical techniques investigated in this thesis.

Two different real data sets were used concerning different problems. One, refers to the problem of unemployment after apprenticeship while the other is a credit–scoring problem.

In the application about unemployment after apprenticeship it turned out that the specification of the parametric model, the logit model, could not be rejected. For the credit-scoring application results are not so clear but one is more inclined to accept also the logit model.

Conclusions

This thesis studies the problem of testing the specification of the link function in a binary response model, the function $F(\bullet)$ in

$$P(Y_i = 1|X_i = x_i) = E(Y_i|X_i = x_i) = F(x_i^T\beta) \quad i = 1, \cdots, n$$

where Y_i can assume only the values 0 or 1.

Chapter 2 introduces the binary response model and explores its natural economic motivation and interpretation. In this framework individuals choose one of the two alternatives by maximizing their indirect utilities which are stochastic. The link function is derived based on the distribution assumptions made for the stochastic utilities of the individual decision makers.

The binary response model can be viewed as a nonlinear regression model where the link function has not to be a distribution function but its range has to be contained in $(0, 1)$ in order that the conditional expectation of the dependent is positive and less than 1.

Classic parametric models define for $F(\bullet)$ a distribution function like the normal, the logistic or the double exponential leading to respectively probit, logit or complementary log-log models. One used strategy assumes that all individuals are identically distributed. In practice this hypothesis may be unrealistic and heterogeneity among individuals should be considered. One can introduce the heterogeneity parametrically by defining a parametric function that models the heteroscedasticity in the model. This is the case of the random-coefficients probit model.

An alternative to parametric models considered in this thesis is the single index model (SIM). This model results from assuming that the link $F(\bullet)$ is unknown but the index, $x_i^T\beta$ is defined parametrically. Thus, this may be considered a semiparametric model. The SIM allows for heterogeneity among individuals if it depends on the index $x_i^T\beta$. Chapter 2 develops the study of the semiparametric binary choice model.

A semiparametric approach has the advantage of avoiding the curse of dimensionality by aggregating a multidimensional variable in a parametric index function. This is the approach taken in this work.

This thesis analyzes and develops techniques that allow to detect unspecified deviations of the link function given a parametric (and particularly linear) specification for the index in a binary choice model. These deviations may include heterogeneity among individuals. The procedure is equivalent to test the specification of a parametric binary choice model against the SIM.

One way is based on the calculation of semiparametric confidence bars and confidence bands for the parametric model. This approach has the advantage of allowing an analysis which is highly graphical and can help on finding the structure of the possible deviation from the assumed model. This procedure is analyzed in chapter 2.

A different strategy assesses the specification of a parametric model by means of a formal specification test. Chapter 3 discusses some specification tests that allow to compare the parametric binary choice model against the semiparametric SIM to conclude that the most convenient procedure is the HH–test introduced by Horowitz and Härdle (1994). The overcoming of the HH–test in this problem is due to a more appealing motivation and to an easier implementation.

The properties about the behavior of the test in models with binary responses were not known. This dissertation studies the performance of the test in finite samples for the binary choice model. This is done in chapter 4. The conclusions are that the test statistic shows a negative bias and a bandwidth dependency under the null which damage severely the power of the test. The negative bias is mainly due to the need of estimating the index to calculate the statistic.

Next, this dissertation proposes an improvement of the HH–test in order to reduce the harmful consequences of the negative bias and bandwidth dependency under the null. It consists in finding more accurate critical values for the test in finite samples. Two approaches are explored, the use of bootstrap and the derivation of analytical corrections.

Chapter 5 proposes a bootstrap procedure of the HH–test, the so–called BHH–test. Results of a simulation study reveal that the bootstrap is clearly beneficial in this problem.

Chapter 6 introduces a modification of the HH–test, called MHH–test, which has the same asymptotical distribution as the first but a better approximation for the first and second moments in finite samples and reduces the bandwidth dependency under the null. This approximation was deduced taking into account the distortion caused by the parametric estimation of the index function necessary to calculate the test statistic. Therefore, using the corrected moments the negative bias is almost eliminated which leads to a improvement in the power of test in finite samples. The simulation studies show that this improvement is very satisfactory.

The improvements of the HH–test defined in this thesis have the counterpart of a greater complexity of calculations. Specially for the MHH–test where the correction of the statistic moments evolves the calculation of products of matrices of size $n \times n$ with n the sample size which is very computational demanding for samples from moderate to big sizes. In this case the use of the bootstrap procedure introduced is specially advised. A comparison between the performance of the BHH–test and the MHH–test has shown that there is not a clear superiority from one of the procedures. If possible both should be calculated.

A relevant and very difficult problem, not completely answered in this thesis, concerns the choice of the bandwidth that maximizes the power of the test keeping the size in the right level. Undersmoothing and oversmoothing have in general a pernicious effect leading to false rejections and consequently should be rejected. Oversmoothing usually leads to great negative values for the HH–statistic. The one–sided test helps to prevent from those false rejections. The one–sided test is supported by the result of Horowitz and Härdle (1994) that under the alternative the statistic has an asymptotic positive mean.

This dissertation give some guidelines about bandwidth choice for the specification tests in study based mainly on empirical considerations. First, it is advised to start with a plot of the semiparametric fit for several bandwidths and then choose the smallest one that does not give a very wiggly curve. It was noticed that the size of the bandwidth is related to the particular variance of the index. It would be interesting to investigate further this issue in order to find an optimal relation between both. This is a subject to be tackled in future work.

The testing procedures studied are more able to detect deviations from the parametric model that have the shape of bumps. This includes also some patterns of heterogeneity. Remember the shape of the logit with heteroscedasticity in Figure 1.8 where the probability curve is plotted against the index divided by the respective heteroscedastic variances. The not so good results obtained for the CLL model are understandable given the closeness between the logit fit and the true CLL. But if the logit curve depicts so well the features of the true model in cases where it is not the logit itself then misspecification is not serious and on the other side one needs a enormously quantity of observations to make a testing procedure being able to reject the misspecified fit.

This thesis also shows how the specification methods analyzed can be successfully implemented using an adequate and comfortable computational environment like XploRe 3.

Appendix A

Proof of Theorem 6.1

The proof of theorem 6.1 follows the proof in Horowitz and Härdle (1994) closely, although fewer approximations are needed for the modified statistic than for the original one. The assumptions, mainly conditions on boundedness, and smoothness can be found in Horowitz and Härdle (1994, pp. 25–26). We begin with a lemma that shows that we can substitute the statistic by an approximation with a deterministic denominator. The asymptotic distribution is then derived from a central limit theorem for degenerate U-statistics of Hall (1984). In addition to the notation introduced in Section 6, we use $\bar{\tilde{r}}(\eta_i) = \tilde{r}(\eta_i)\hat{p}^h(\eta_i)/p(\eta_i)$, with $\eta_i = x_i^T\beta$ and $p(\eta_i)$, the "true" density of the linear predictors, and its estimate

$$\hat{p}^h(\eta_i) = \frac{1}{nh}\sum_{j\neq i} K[(\eta_j - \eta_i)/h].$$

Lemma A.1 *Define* $MT_1^h = \sqrt{h}\sum_i u(\eta_i)r(\eta_i)\bar{\tilde{r}}(\eta_i)$. *Under the assumptions of Horowitz and Härdle (1994)*

$$MT^h = MT_1^h + o_p(1)$$

Proof:
Notice that $MT^h = MT_1^h + R_1$ with $R_1 = \sqrt{h}\sum_i u(\eta_i)[\tilde{r}(\eta_i) - \bar{\tilde{r}}(\eta_i)]$. By lemma 1 of Horowitz and Härdle (1994) $|p(\eta) - \hat{p}^h(\eta)| = O((\log n)^{1/2}/(nh)^{1/2})$ almost surely and uniformly over η. That allows us to write

$$\tilde{r}(\eta_i) - \bar{\tilde{r}}(\eta_i) = \frac{\bar{\tilde{r}}(\eta_i)[p(\eta_i) - \hat{p}^h(\eta_i)]}{p(\eta_i)} + O((\log n)^{1/2}/(nh)^{1/2})$$

almost surely and $R_1 = \sqrt{h}\sum_i u(\eta_i)r(\eta_i)\bar{\bar{r}}(\eta_i)[p(\eta_i) - \hat{p}^h(\eta_i)]/p(\eta_i) + o_p(1)$. $E(R_1) = 0$ because $\bar{\bar{r}}(\eta_i)$ does not depend on $r(\eta_i)$. Next we shall prove that $V(R_1) = o(1)$ where $V(\bullet)$ stands for variance.

$$\begin{aligned} V(R_1) &= hE\left\{\sum_i u^2(\eta_i)r^2(\eta_i)\bar{\bar{r}}(\eta_i)^2[p(\eta_i) - p^h(\eta_i)]^2/p(\eta_i)^2\right. \\ &+ \sum_i\sum_{j\neq i} u(\eta_i)u(\eta_j)r(\eta_i)r(\eta_j)\bar{\bar{r}}(\eta_i)\bar{\bar{r}}(\eta_j) \\ &\left.[p(\eta_i) - \hat{p}^h(\eta_i)][p(\eta_j) - \hat{p}^h(\eta_j)]/p(\eta_i)p(\eta_j)\right\} \end{aligned}$$

and

$$\begin{aligned} &E\{u(\eta_i)r(\eta_i)\bar{\bar{r}}(\eta_j)[p(\eta_j) - \hat{p}^h(\eta_j)]/p(\eta_i)\} \\ &\quad = u(\eta_i)\sigma^2(\eta_i)E\{K[(\eta_i - \eta_j)/h](1/nh)[1/p(\eta_j)^2][p(\eta_j) - \hat{p}^h(\eta_j)]\} \\ &\quad = o(n^{-1}) \end{aligned}$$

Therefore,

$$\begin{aligned} V(R_1) &= hE\left\{u(\eta_i)^2r(\eta_i)^2\bar{\bar{r}}(\eta_i)^2[p(\eta_i) - \hat{p}^h(\eta_i)]^2/p(\eta_i)^2\right\} \\ &\leq [h/p(\eta_i)^2]E\left\{\sum_i u(\eta_i)^2r(\eta_i)^2\bar{\bar{r}}(\eta_i)^2O((\log n)/(nh))\right\} \end{aligned}$$

almost surely. Following Bierens (1987) (pp 104) $E[\bar{\bar{r}}(\eta_i)^2] = O(1/(nh))$ uniformly over η (and given that by assumption $u(\bullet)$ is bounded) consequently $V(R_1) = o(1)$. From Chebyshev's inequality $R_1 = o_p(1)$. Q.E.D.

Proof of theorem 1:

Define $Z_i = (\eta_i, r(\eta_i))$ and

$$\begin{aligned} A_n(Z_i, Z_j) &= \frac{1}{\sqrt{h}n}u(\eta_i)r(\eta_i)r(\eta_j)\frac{K[(\eta_j - \eta_i)/h]}{p(\eta_i)} \\ H_n(Z_i, Z_j) &= A_n(Z_i, Z_j) + A_n(Z_j, Z_i) \\ &= \frac{1}{\sqrt{h}n}r(\eta_i)r(\eta_j)\left[\frac{u(\eta_i)}{p(\eta_i)} + \frac{u(\eta_j)}{p(\eta_j)}\right]K[(\eta_i - \eta_j)/h] \\ \mu(Z_i) &= E[H_n(Z_i, Z_j)|Z_i] = 0 \\ Q_n(Z_i, Z_j) &= E[H_n(Z_l, Z_i)H_n(Z_l, Z_j)|Z_i, Z_j] \end{aligned}$$

Given Lemma 1 and after some algebra the modified HH–statistic can be written

$$MT^h = \sum_i\sum_{j<i} H_n(Z_i, Z_j) + o_p(1)$$

The first term in the formula above is a U-statistic and given that $\mu(Z_i) = 0$ this is a degenerate U-statistic. According to Theorem 1 of Hall (1984) this U-statistic will converge in distribution to a $N(0, \sigma_T^2)$ if

$$\{E[Q_n(Z_i, Z_j)^2] - n^{-1}E[H_n(Z_i, Z_j)^4]\}/\{E[H_n(Z_i, Z_j)^2]\}^2 \to 0, \qquad (A.1)$$

$$(1/2)n^2E[H_n(Z_i, Z_j)^2] \to \sigma_T^2 \qquad (A.2)$$

as $n \to \infty$

$$\begin{aligned} Q_n(Z_i, Z_j) &= \frac{1}{n^2h}E_{l|i,j}\left\{r(\eta_l)r(\eta_i)\left[\frac{u(\eta_l)}{p(\eta_l)} + \frac{u(\eta_i}{p(\eta_i)}\right]\right. \\ &\times \left. K[(\eta_i - \eta_l)/h]K[(\eta_j - \eta_l)/h]r(\eta_l)r(\eta_i)\left[\frac{u(\eta_l)}{p(\eta_l)} + \frac{u(\eta_i}{p(\eta_i)}\right]\right\} \\ &= \frac{1}{n^2h}\int r(\eta_i)r(\eta_j)R(\eta_l, \eta_i, \eta_j)K[(\eta_i - \eta_l)/h]K[(\eta_j - \eta_l)/h] \end{aligned}$$

$p(\eta_l)d\eta_l$

and $R(\bullet)$ is a bounded continuous function. Making a change of variable $\xi = (\eta_l - \eta_i)/h$ gives,

$$Q_n(Z_i, Z_j) = \int r(h\xi + \eta_i, \eta_i, \eta_j)K(\xi)K[\xi + (\eta_i - \eta_j)/h]p(\eta_i + h\xi)d\xi$$

and

$$\begin{aligned} E[Q_n(Z_i, Z_j)^2] &= cn^{-4}\int R_n(h\xi_1 + \eta_i, h\xi_2 + \eta_j, \eta_i, \eta_j)K(\xi_1)K(\xi_2) \\ &\times K[\xi_1 + (\eta_i - \eta_j)/h]K[\xi_2 + (\eta_i - \eta_j)/h]p(h\xi_1 + \eta_i) \\ &\times p(\xi_2 + \eta_j)p(\eta_i)p(\eta_j)d\xi_1 d\xi_2 d\eta_i d\eta_j \\ &= O(n^{-4})O(h) = o(1) \qquad (A.3) \end{aligned}$$

Regarding $E[H_n(\bullet)^4]$ note that,

$$H_n(Z_i, Z_j)^4 = n^{-4}h^{-2}\bar{R}_n(Z_i, Z_j)$$

and $\bar{R}_n(Z_i, Z_j)$ is bounded uniformly over n. Therefore

$$n^{-1}E[H_n(Z_i, Z_j)^4] = O(1/(n^5h^2)) \qquad (A.4)$$

Considering $E[H_n(\bullet)^2]$ we have

$$\begin{aligned} H_n(Z_i, Z_j)^2 &= A_n(Z_i, Z_j)^2 + A_n(Z_j, Z_i)^2 + 2A_n(Z_i, Z_j)A_n(Z_j, Z_i) \\ A_n(Z_i, Z_j)^2 &= \frac{1}{n^2hp(\eta_i)^2}u(\eta_i)^2r(\eta_i)^2r(\eta_j)^2K[(\eta_j - \eta_i)/h]^2 \\ n^2E[A_n(Z_i, Z_j)^2] &= \int \frac{u(\eta_i)^2}{p(\eta_i)^2}\sigma^2(\eta_i)\sigma^2(h\xi + \eta_i)K(\xi)^2p(\eta_i)p(h\xi + \eta_i)d\xi d\eta_i \\ &= C_K\int u(\eta)^2[\sigma^2(\eta)]^2d\eta + o(1) \qquad (A.5) \end{aligned}$$

By arguments similar to those used to obtain (A.5)

$$n^2 E[A_n(Z_i, Z_j)A_n(Z_j, Z_i)] = C_K \int u(\eta)^2[\sigma^2(\eta)]^2 d\eta + o(1)$$

Therefore,

$$n^2 E[H_n(Z_i, Z_j)^2] = 4C_K \int u(\eta)^2[\sigma^2(\eta)]^2 d\eta + o(1) \tag{A.6}$$

Given results (A.3), (A.4) and (A.6) conditions (A.1) and (A.2) are verified. Q.E.D.

Appendix B

XploRe Procedures

```
; ***********************************************************************
; * UNIFBAND macro ******************************************************
; ***********************************************************************
; * Function: UNIFBAND estimates semiparametric uniform confidence band *
; *           at a% level for a binary responses model.                 *
; *                                                                     *
; * Call:     y = UNIFBAND (x y h {a {m}})                              *
; *                                                                     *
; * ->X       x  n x 1  matrix   projected index                        *
; *           y  n x 2  matrix   1. col = dependent variable (0 or 1)   *
; *                              2. col = parametric fitted P(Y=1|X)    *
; *           h         scalar   bandwidth for kernel regression        *
; *           a         scalar   0 < a < 100 ; default=90               *
; *           m  n x 1           m should be given only If the data is  *
; *                              binomial with m the vector with the    *
; *                              binomial coeficients.                  *
; *                                                                     *
; * X->       c  n x 4  matrix   1. col = grid where the C.L. are calc. *
; *                              2. col = semiparametric est. P(Y=1|X)  *
; *                              3. col = lower confidence limit        *
; *                              4. col = upper confidence limit        *
; *                                                                     *
; * Example:  LIBRARY(GLM)                                              *
; *           x = read(kyphosis)                                        *
; *           y = x[,4]                                                 *
; *           x = x[,1:3]                                               *
; *           x = matrix(rows(x))~x                                     *
; *           (itres b bvar stat) = GLMBILO(x y)                        *
```

```
; *           v  = x*b                                          *
; *           z  = y~itres[,1]                                  *
; *           h  = 2                                            *
; *           LIBRARY(SMOOTHER)                                 *
; *           LIBRARY(ADDMOD)                                   *
; *           cb = unifband(v z h)                              *
; *                                                             *
; *************************************************************
; ** Isabel Proenca, 940727 **********************************
; *************************************************************
proc(cb) = unifband(x y h a m)
  vex = 0
  error(cols(x).<>1 "UNIFBAND: COLS(X) <> 1")
  error(cols(y).<>2 "UNIFBAND: COLS(y) <> 2")
  if((exist(a).=1))
      error(cols(a).<>1 "UNIFBAND: COLS(a) <> 1")
      error(a.<0 .| a .>= 100  "UNIFBAND: wrong choice of a")
  else
      a = 90
  endif
  if((exist(m).=1))
      error(cols(m).<>1 "UNIFBAND: COLS(m) <> 1")
  else
      m = 1
  endif
   ct   = -log(-0.5*log(a/100))
   k    = cols(x)
   n    = rows(x)
   yhat = y[,2]
   y    = y[,1]
   yhat = sort(x~yhat 1)
   x    = sort(x~y 1)
   y    = x[,2]
   x    = x[,1]
   delt = 4/3
   h    = h^delt
   s    = h*n^(0.8/5)  ; bandwidth for bias correction
   br   = (h/s)^2
   (num den) = sker(x y s)
   ms = (num./den)
   (num den) = sker(x y h)
   mh = num./den
   fhat  = (mh - br*ms)./(1-br) ; Bierens' bias correction
   ;fhat = mh
```

```
    fhat = fhat.*(fhat.>0).*(fhat.<=1) + (fhat.>1)
    vy = m.*mh.*(1-mh)
    ck = 5/7
    c1 = sqrt(0.4*delt*log(n))
    dn = c1 + (1/c1)*0.5*log(3./(4*pi^2))
    cbw = (ct./c1 + dn)*sqrt((vy.*ck)./den)
    cbl = fhat - cbw
    cbl = cbl.*(cbl.>0)
    cbu = fhat + cbw
    cbu = cbu.*(cbu .<= 1) + (cbu.>1)
    cb  = x~fhat~cbl~cbu
; show the parametric regression  and the confidence bands
   cbu  = x~cbu~mask(n 1 white)
   cbl  = x~cbl~mask(n 1 white)
   npt  = ceil(n/10);
   q    = ((aseq(0 n 1)%npt).=0)
   fhat = x~fhat~vtocc((q.=0).*31 + (q.=1).*95)
   yhat = yhat~mask(n 1 yellow)
   y    = x~y~mask(rows(y) 1 X white)
   tit  = string("uniform c.b. at %2.0f%%" a)
   capture on
   writecon(27)
   show(yhat cbl cbu fhat y s2d)
   u = update(y 5 point)
   u = update(yhat 1 solid line thick xaxis "index ")
   u = update(cbl 2 line dotted title tit)
   u = update(cbu 3 line dotted yaxis "E(Y|X)")
   u = update(fhat 4 line solid)
   show(yhat cbl cbu fhat y s2d)
endp

; **********************************************************************
; * HHTEST macro ********************************************************
; **********************************************************************
; * Function: HHTEST calculates the H-H statistic to test the specifi-  *
; *                  cation of the link function  of a  binary  choice  *
; *                  model (assuming the index is correctly specified). *
; *                                                                     *
; * Call:     (t  p) = HHTEST (vhat y yhat h {c {m}})                   *
; *                                                                     *
; * ->X       vhat  n x 1  matrix with the projected index              *
; *           y     n x 1  matrix                                       *
; *           yhat  n x 1  matrix with the parametric estimate of E(Y|X)*
; *           h     scalar (positive) -- the bandwidth for kernel       *
```

```
; *                                     regression with Quartic kernel *
; *            c      scalar  0 =< c < 1 (optional)  -- proportion of the *
; *                   sample to be cut in each extreme. Default is 0.05.  *
; *            m      n x 1  or the scalar 1. m should be given only for  *
; *                   binary responses. If the data is binomial m is the  *
; *                   vector with the binomial coeficients. If the data is*
; *                   bernouli, m=1. m is necessary to calculate the vari-*
; *                   ance of y. If y is not binary the variance will be  *
; *                   given by a nonparametric regression of  (y-fhat)^2  *
; *                   on vhat.                                            *
; *                                                                       *
; * X->        t      scalar  -- the statistic value                      *
; *            p      scalar  -- the p-value of t                         *
; *                                                                       *
; * Example:   LIBRARY(GLM)                                               *
; *            x = read(kyphosis)                                         *
; *            y = x[,4]                                                  *
; *            x = x[,1:3]                                                *
; *            x = matrix(rows(x))~x                                      *
; *            h = 2                                                      *
; *            (itres beta bvar stat) = GLMBILO(x y)                      *
; *            LIBRARY(ADDMOD)                                            *
; *            LIBRARY(SMOOTHER)                                          *
; *            (t  p) = hhtest(itres[,2] y itres[,1] h 0.05 1)            *
; *                                                                       *
; * Comments: "Testing a parametric model against a semiparametric        *
; *             alternative", J. Horowitz and W. Haerdle, Econometric     *
; *             Theory, forthcoming.                                      *
; *                                                                       *
; ************************************************************************
; ** Isabel Proenca, 940724 **********************************************
; ************************************************************************
proc (t p)= hhtest(vhat y yhat h c m)
  error(cols(y)<>1 ": COLS of Y <> 1")
  error(cols(yhat)<>1 ": COLS of YHAT <> 1")
  error(cols(vhat)<>1 ": COLS of VHAT <> 1")
;
  capture on
  x = vhat~yhat~y
  n = rows(x)
  x = sort(x 1)
  vhat = x[,1]
  yhat = x[,2]
  y    = x[,3]
```

```
  free(x)
  res1 =  y - yhat
  s = h*n^(0.8/5)
  br = (h/s)^2
  IF ((exist(c).=0))
     c = 0.05        ; proportion to cut off in each side of the sample
  else
    error((c.<0) .| (c.>=1) ": wrong choice of c: 0 =< c < 1")
  ENDIF
;
  (fhath den) = krgloo(vhat y h)
  (fhats den) = krgloo(vhat y s)
  free(den)
  fhat  = (fhath - br*fhats)./(1-br) ; remove bias from fhat
  mhi   = vhat~fhat
  res2  = fhat-yhat
;
; cut off the extreme values
;
  ind1 = ceil(c*n)+1
  ind2 = ceil((1-c)*n)
  res1  = res1[ind1:ind2,]
  res2  = res2[ind1:ind2,]
;
; conditional variance of Y
;
  if((exist(m).=1))
     error(cols(m)<>1 .| (m<>1) ": wrong choice of m")
     vy = m.*fhat.*(1-fhat)
  else
     rs = (y - fhat).*(y - fhat)
     (num den) = sker(vhat rs h)
     vy = num./den
  endif
  vy   = vy[ind1:ind2,]
;
; H-H test statistic
;
  t = sum(res1.*res2)
;
; variance of t
;
  ck = 5/7
  (fu phat) = sker(vhat 0 h)
```

```
  phat = phat[ind1:ind2,]
  itg = sum(vy^2./phat)
  vt = 2*ck*itg
  t = t/sqrt(vt)
  p = cdfn(t)
;
; plot of parametric and semiparametric estimates
;
  par = vhat~yhat
  dat = vhat~y
  dat = dat~mask(n 1 X white)
  writecon(27)
  show(dat par mhi s2d)
  u = update(par 2 yaxis "E(Y|X)")
  u = update(par 2 line solid xaxis "index  ")
  u = update(mhi 3 line solid title "parametric and LOO kernel")
  writecon(27)
  show(dat par mhi s2d)
  mh = regest(vhat~y h)
  writecon(27)
  show(dat par mh s2d)
  u = update(par 2  yaxis "E(Y|X)")
  u = update(par 2 line solid xaxis "index  ")
  u = update(mh  3 line solid title "parametric and kernel estim.")
  writecon(27)
  show(dat par mh s2d)
endp

; ***********************************************************************
; * MODHHTST macro********************************************************
; ***********************************************************************
; * Function: MODHHTST calculates the modified HH-statistic to test the *
; *           specification of the link function  of a  binary  choice  *
; *                  model (assuming the index is correctly specified). *
; *                                                                     *
; * Call:     (t  p bc) = MODHHTST (x vhat y yhat h)                    *
; *                                                                     *
; * ->X       x     n X k  matrix with the explanatory variables        *
; *           vhat  n X 1  matrix with the projected index              *
; *           y     n x 1  matrix                                       *
; *           yhat  n x 1  matrix with the parametric estimate of E(Y|X)*
; *           h     scalar (positive) -- the bandwidth for kernel       *
; *                                 regression with Quartic kernel *
; * X->       t     scalar  -- the statistic value                      *
```

```
; *           p     scalar  -- the p-value of t                        *
; *                                                                    *
; * Example:  LIBRARY(GLM)                                             *
; *           x = read(sim)                                            *
; *           y = x[,3]                                                *
; *           x = x[,1:2]                                              *
; *           (itres beta bvar stat) = GLMBELO(x y 0)
; *              LIBRARY(SMOOTHER)
; *              LIBRARY(ADDMOD)          *
; *           (t  p bc) = hhtest(x itres[,1] y itres[,2] h)            *
; *                                                                    *
; * Comments: "
; *                                                                    *
; **********************************************************************
; ** Isabel Proenca, 940627 ********************************************
; **********************************************************************
proc (t p) = modhhtst(x vhat y yhat h)
  error(cols(y)<>1 ": COLS of Y <> 1")
  error(cols(yhat)<>1 ": COLS of YHAT <> 1")
  error(cols(vhat)<>1 ": COLS of VHAT <> 1")
  capture on
  dat = vhat~yhat~y~x
  n = rows(dat)
  dat = sort(dat 1)
  x    = dat[,4:cols(dat)]
  vhat = dat[,1]
  yhat = dat[,2]
  y    = dat[,3]
  free(dat)
  res =  y - yhat
;
; H matrix
;
  vmat = yhat.*(1.-yhat);
  vx   = vmat.* x;
  a    = inv((x'* vx));
  hmat = vx*a*x';
  hmat = unit(n) - hmat;
  free vx;
;
; smoothing matrix with weight fc
;
  dif  = (vhat .- vhat')./h
  wmat = quartic(dif)
```

```
  free(dif)
  q0   = quartic(0)
  wmat = wmat .- diag(matrix(n 1 q0));
  den  = sum(wmat) .- q0;
  ind  = (den .= 0);
  wmat = wmat./(den .+ ind);
  wmat = wmat.*(1.-ind);
  free (den ind);
;
; calculate the statistic terms
;
   resmo = wmat*res;
   tn    = res'*resmo;
;
; calculate the variance and the bias correction
;
   dmat = hmat'*wmat*hmat;
   free wmat;
   par1 = 0;
   bc = 0;
   i = 1;
   while (i .<= n);
           cold  = dmat[,i];
           si    = vmat[i,];
           bc      = bc + cold[i,]*si;
           cold    = cold.*cold;
           par2    = cold[i,]*(si-4*si^2);
           cold[i,] = 0;
           par1      = par1+2*si*sum(cold.*vmat) + par2;
           i = i+1;
   endo;
   free (cold dmat);
   st   = sqrt(par1);
   tn   = tn - bc;
   t    = tn./st;
   write(t t)
   p    = pdfn(t)
   write(p p)
;
; plot of parametric and semiparametric estimates
;
  res = vhat~res
  resmo =  vhat~resmo
  zero = (vhat[1,]|vhat[n,])~(0|0)
```

```
  ;dat = dat~mask(n 1 X white)
  writecon(27)
  show(zero  res resmo s2d)
  u = update(zero 1 line dashed  yaxis "residuals")
  u = update(res 2 point xaxis "index  ")
  u = update(resmo 3 line solid title "residuals -- LOO kernel")
  writecon(27)
  show(zero res resmo  s2d)
  mh = regest(res h)
  writecon(27)
  show(zero res mh s2d)
  u = update(zero 1 line dotted  yaxis "residuals")
  u = update(res 2 point xaxis "index  ")
  u = update(mh  3 line solid title "residuals -- kernel estim.")
  writecon(27)
  show(zero res mh s2d)
endp
```

Bibliography

Amemiya, T. (1981). Qualitative response models: a survey, *Journal of Economic Literature* **XIX**: 1483–1536.

Amemyia, T. (1983). Nonlinear regression models, *in* Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, North Holland, Amsterdam, pp. 333–390.

Azzalini, A., Bowman, A. and Härdle, W. (1989). On the use of nonparametric regression for model checking, *Biometrika* **76**: 1–11.

Berridge, D. M. (1993). Modelling ordinal recurrent events. Unpublished manuscript.

Bierens, H. J. (1987). Kernel estimators of regression functions, *in* T. F. Bewley (ed.), *Advances in Econometrics: Fifth world congress*, Vol. 1, Cambridge University Press, New York, pp. 99–144.

Bierens, H. J. (1990). A consistent conditional moment test of functional form, *Econometrica* **58**: 1443–1458.

Bonneu, M., Delecroix, M. and Malin, E. (1993). Semiparametric versus nonparametric estimation in single index regression model: A computational approach, *Computational Statistics* (8): 207–222.

Chappell, R., Czado, C. and Newton, M. A. (1993). Semiparametric bayesian inference for binary regression. Unpublished manuscript.

Chesher, A. (1984). Testing for neglected heterogeneity, *Econometrica* **52**: 865–872.

Collet, D. (1991). *Modelling Binary Data*, Chapman and Hall, London.

Czado, C. (1992). On link selection in generalised linear models, *in* L. Fahrmeir, B. Francis, R. Gilchrist and G. Tutz (eds), *Advances in GLIM and Statistical Modelling. Proceedings of the the GLIM92 Conference and the 7th*

International Workshop on Statistical Modelling, Springer–Verlag, New York, pp. 60–65.

Delecroix, M. and Hristache, M. (1994). Convergence d'une classe de m–estimateurs semi–paramétriques dans les modeles à direction révélatrice unique. Unpublished manuscript.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, manuscript.

Firth, D. (1992). Bias reduction, the jeffreys prior and glim, *in* L. Fahrmeir, B. Francis, R. Gilchrist and G. Tutz (eds), *Advances in GLIM and Statistical Modelling. Proceedings of the the GLIM92 Conference and the 7th International Workshop on Statistical Modelling*, Springer-Verlag, New York, pp. 91–100.

Gozalo, P. L. (1993). A consistent model specification test for nonparametric estimation of regression function models, *Econometric Theory* **9**: 451–477.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of multivariate analysis* **14**(1): 1–16.

Hall, P. (1992). *The Bootstrap and the Edgeworth Expansion*, Springer, New York.

Hall, P. and Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of monte carlo tests, *Journal of the Royal Statistic Society, Series B* **51**(3): 459–467.

Härdle, W. (1990). *Applied Nonparametric Regression*, Econometric Society Monographs, 19, Cambridge University Press.

Härdle, W. and Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands, *Journal of the American Statistical Association* **83**: 102–110.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric *versus* parametric regression fits, *Annals of Statistics* **21**(4): 1926–1947.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression, *Annals of Statistics* **19**: 778–796.

Härdle, W. and Stoker (1989). Investigating smooth multiple regression by the method of average derivatives, *Journal of the American Statistical Association* **84**: 986–995.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.

Hausman, J. A. and Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogenous preferences, *Econometrica* **46**: 403–426.

Hildenbrand, W., Härdle, W. and Jerison, M. (1991). Empirical evidence on the law of demand, *Econometrica* **59**: 1525–1549.

Horowitz, J. and Härdle, W. (1994). Testing a parametric model against a semiparametric alternative, *Econometric theory* **forthcoming**.

Horowitz, J. L. (1993). Semiparametric estimation of a work-trip mode choice model, *Journal of Econometrics* **58**(1/2): 49–70.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models, *Journal of Econometrics* **58**(1/2): 71–120.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. (1985). *The Theory and Practice of Econometrics*, Wiley, New York.

Klein, R. and Spady, R. (1993). An efficient semiparametric estimator for binary response models, *Econometrica* **61**(2): 387–421.

le Cessie, S. and van Houwelingen, J. C. (1991). A goodness-of-fit test for binary regression models based on smoothing methods, *Biometrics* **47**: 1267–1282.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

McFadden, D. (1981). Econometric models of probabilistic choice, *in* C. F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Mass., pp. 198–272.

Mosbach, O. (1992). One–step bootstrapping in generalized linear models, *in* K.-H. Jöckel, G. Rothe and W. Sendler (eds), *Bootstrap and related techniques. Proceedings of an International Conference held in Trier, FRG, 1990*, Springer-Verlag, New York, pp. 143–147.

Nadaraya, E. A. (1964). On estimation regression, *Theory of Probability and its Applications* **9**: 141–142.

Newey, W. (1985). Maximum likelihood specification testing and conditional moment test, *Econometrica* **53**: 1047–1070.

Powell, J., Stock, J. and Stoker, T. (1989). Semiparametric estimation of index coefficients, *Econometrica* **57**(6): 1403–1430.

Proença, I. (1993). On the performance of the H-H test, *Discussion Paper 9305*, CORE, Université Catholique de Louvain.

Proença, I. and Ritter, C. (1993). Negative bias in the H-H statistic, *Discussion Paper 9321*, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

Proença, I. and Ritter, C. (1994). Semiparametric testing of the link function in models for binary outcomes, *Discussion Paper 9415*, Institut de Statistique, Université Catholique de Louvain.

Proença, I. and Werwatz, A. (1994). Comparing parametric and semiparametric binary response models, *in* W. Härdle, S. Klinke and B. A. Turlach (eds), *XploRe 3 — An interactive statistical computing environment*, Springer Verlag, chapter 14.

Projektgruppe (1991). Das sozio-ökonomische panel, *Das Sozio-ökonomische Panel (SOEP) im Jahre 1990/91*, Deutches Institut für Wirtschaftsforschung. Vierteljahreshefte zur Wirtschaftsforschung, pp. 146–155.

Rodríguez-Campos, M. C. and Cao-Abad, R. (1983). Nonparametric bootstrap confidence intervals for discrete region functions, *Journal of Econometrics* **58**: 207–222.

Ruud, P. A. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution, *Econometrica* **51**: 225–228.

Ruud, P. A. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution, *Journal of Econometrics* **23**: 157–187.

Serfling, R. J. (1985). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Stoker, T. M. (1992). *Lectures on Semiparametric Econometrics*, CORE Lecture Series, Louvain-La-Neuve.

Train, K. (1986). *Qualitative Choice Analysis. Theory, Econometrics and an Application to Automobile Demand*, The MIT Press, London.

Turlach, B. (1994). *Computer-aided Additive Modeling*, PhD thesis, Institut de Statistique, Université Catholique de Louvain, Louvain-La-Neuve.

van Soest, A. (1992). Discrete choice models of family labour supply, *Discussion Paper 9214*, CentER, Tilburg University, The Netherlands.

Watson, G. S. (1964). Smooth regression analysis, *Sankhyā, Series A* **26**: 359–372.

Whang, Y.-J. and Andrews, D. W. K. (1993). Tests of specification for parametric and semiparametric models, *Journal of Econometrics* **57**: 277–318.

White, H. (1982). Maximum likelihood estimation of misspelled models, *Econometrica* **50**: 1–25.

PUBLICATIONS DE LA FACULTE DES SCIENCES ECONOMIQUES, SOCIALES ET POLITIQUES DE L'UNIVERSITE CATHOLIQUE DE LOUVAIN

Nouvelle série.

1. BARDOS-FELTORONYI N., Diffusion du progrès et convergence des prix (Allemagne-Angleterre 1792-1913). Louvain, Editions Nauwelaerts; Paris, Béatrice-Nauwelaerts, 1966. Un vol. in-8° de 230 pp.

2. SZAPARY G., Diffusion du progrès et convergence des prix (Europe - Etats-Unis 1899-1962). Louvain, Editions Nauwelaerts; Paris, Béatrice-Nauwelaerts, 1966. Un vol. in-8° de 246 pp.

3. DELEECK H., Maatschappelijke zekerheid en inkomensherverdeling in Belgïe. Leuven, Standaart Wetenschappelijke Uitgeverij, 1966. Un vol. in-8° de 347 pp.

4. VAN ACOLEYEN M. Bedrijfseconomische aspecten van de produktiviteits analyse. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1966. Un vol. in-8° de 295 pp.

5. DE VYLDER F., Processus de Poisson généralisés. Louvain, Librairie Universitaire, 1966. Un vol. in-8° de 187 pp.

6. DOBBELAERE K., Sociologische analyse van de katholiciteit. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1966.Un vol. in-8° de 413 pp.

7. REMY J., La ville : phénomène économique. Bruxelles, Les Editions ouvrières, 1966. Un vol. in-8° de 297 pp.

8. PRADES J.A., La sociologie de la religion chez Max Weber, essai d'analyse et de critique de la méthode. Louvain, Editions Nauwelaerts, Paris, Béatrice-Nauwelaerts, 1966. Un vol. in-8° de 293 pp.

9. ROMANIUK A., La fécondité des populations congolaises. Paris, Editions Mouton & Cie, 1967. Un vol. in-8° de 348 pp.

10. EL-BAKJAJI S.el-Din., Processus stochastiques, applicables à l'étude des Caisses de Pensions, Processus de Markov, Processus de renouvellement. Louvain, Librairie Universitaire, 1967. Un vol. in-8° de 113 pp.

11. BUZZI A.R., La théorie politique d'Antonio Gramsci. Louvain, Editions Nauwelaerts; Paris, Béatrice-Nauwelaerts, 1967. Un vol. in-8° de 363 pp.

12. DE DECKER H., Développement communautaire, une stratégie d'édification de la nation. Analyse des modèles de développement communautaire en Guinée et au Sénégal. Kinshasa, Editions de l'IRES, Lovanium; Paris, Mouton & Cie, 1967. Un vol. in-8° de 470 pp.

13. VANNESTE O., Het groeipool concept en de regionaal-economische politiek, toepassing op de Westvlaamse economie. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967. Un vol. in-8° de 376 pp.

14. HICKEY J., Urban Catholics in England and Wales : from 1829 to the present day. Londres, Geoffrey Chapman ltd., 1967. Un vol. in-8° de 188 pp.

15. DELIEGE-ROTT D., Le Médecin face au marché commun. Louvain, Librairie Universitaire, 1967. Un vol. in-8° de 418 pp.

16. CORTEN A., Valeurs sociales et économies au seuil de la croissance. Essai de sociologie de la connaissance. Louvain, Editions Nauwelaerts, 1967. Un vol. in-8° de 235 pp.

17. VAN WAAS O., La pharmacie d'officine face à son avenir - quelques grands problèmes actuels. Louvain, 1967. Un vol. in-8° de 240 pp.

18. DE WACHTER W., De wetgevende verkiezingen als proces van machtverwering in het Belgisch politiek bestel. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967. Un vol. in-8° de 392 pp.

19. MAES R., De overheidsbewegingen op het gebied van de electriciteitsvoorziening in Belgïe. Brugge, Die Keure, 1967. Un vol. in-8° de 531 pp.

20. DELBEKE H., De budgetering in de ondernemingsorganisatie, een critische studie. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967. Un vol. in-8° de 492 pp.

21. DENDOOVEN A., Ontstaan, structuur en werking van de vlaamse K.A.J.. Kessel-Lo, van der Poorten, 1967. Un vol. in-8° de 484 pp.

22. TRAPPENIERS F., Les avantages comparatifs dans le marché commun européen. Louvain, Editions Nauwelaerts, 1967. Un vol. in-8° de 346 pp.

23. BOKONGA C., Tendances fondamentales de l'évolution des relations entre employeurs, travailleurs et gouvernements en Afrique. Louvain, Imprimerie Offset Frankie, 1967. Un vol. in-8° de 185 pp.

24. MANDY P., Progrès technique et Emploi. Aspects structurels de la Théorie de l'Emploi. Louvain, Librairie Universitaire, 1967. Un vol. in-8° de 336 pp.

25. ORBAN E., Le Conseil législatif de Québec 1867-1967. Louvain, 1967. Un vol. in-8° de 354 pp.

26. TAVERNIER K., Investeringsdeterminanten in de particuliere nijverheid, een theoretische benadering met empirische toetsing voor Belgische bedrijfstakken. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967 . Un vol. in-8° de 259 pp.

27. JACQUEMIN A., L'entreprise et son pouvoir de marché. Louvain, Librairie Universitaire, 1967. Un vol. in-8° de 240 pp.

28. WUNSCH G., Les mesures de la natalité. Quelques applications à la Belgique. Louvain, Département de Démographie, 1967. Un vol. in-8° de 146 pp.

29. VAN HOUTVINCK J., Verschillen in de huwelijksvruchtbaarheid van echtparen met onderscheiden beroepsniveau. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967. Un vol. in-8° de 288 pp.

30. DE PILLECYN P.H., Sociaal probleem en verhalend proza. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967. Un vol. in-8° de 195 pp.

31. NGUYEN-ANH-TUAN., Les forces politiques au Sud-Viet-Nam depuis les accords de Genève 1954. Louvain, Librairie Universitaire, 1967. Un vol. in-8° de 362 pp.

32. BRAUERS K., Input-Output analyse en internationale economische integratie. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1968. Un vol. in-8° de 276 pp.

33. CLAEYS-VANHAGENDOREN 25 jaar Belgisch Socialisme. Evolutie van de verhouding van de Belgische werklieden partij tot de parlementaire democratie in Belgïe van 1914 tot 1940. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1967. Un vol. in-8° de 509 pp.

34. TULKENS H., Programming Analysis of the Postal Service : A study in public enterprise economics. Louvain, Librairie Universitaire, 1968. Un vol. in-8° de 316 pp.

35. BOTANA N., La légitimité, problème politique. Buenos-Aires, C.L. Brousse, 1968. Un vol. in-8° de 261 pp.

36. PRESVELOU C., La consommation de la famille : phénomène sociologique. Essai d'une sociologie de la consommation familiale. Bruxelles 19, Les Editions Vie Ouvrière, 1968. Un vol. in-8° de 319 pp.

37. D'HOOGE L., Studie van de heterogeneïteit van een tariefklasse in nietlevensverzekering. Leuven, Imprimerie A. De Wallens, 1968. Un vol. in-8° de 176 pp.

38. PIEL J., Analyse des formes prises par la socialité des adolescents dans le cadre de leurs loisirs. Bruxelles, la Renaissance du Livre, 1968. Un vol. in-8° de 333 pp.

39. EFFANDI A., Tarification du risque individuel en assurance automobile. Louvain, Librairie Universitaire, 1968. Un vol. in-8° de 99 pp.

40. JASKOLD-GABSZEWICZ J., Coeurs et allocations concurrentielles dans des économies d'échanges avec un continu de biens. Louvain, Librairie Universitaire, 1968. Un vol. in-8° de 70 pp.

41. PEEMANS J.-P., Diffusion du progrès économique et convergence des prix. Le cas Congo-Belgique, 1900-1960. La formation du système des prix et salaires dans une économie dualiste. Louvain, Editions Nauwelaerts, 1968. Un vol. in-8° de 520 pp.

42. CHULLIKAL A., Economic Progress and Price Behaviour India-United Kingdom 1860-1966. Louvain, Imprimerie Offset Frankie, 1968. Un vol. in-8° de 251 pp.

43. SLOOF F.L., Het delen van de werknemers in de vermogensaanwas van de produktiegemeenschappen. Alpen aan de Rijn. Uitgeverij N. Samson N.V., 1968. Un vol. in-8° de 227 pp.

44. ROUSSAKIS E., Friedrich List, the Zollverein and the Uniting of Europe. Bruges, College of Europe Press, 1968. Un vol. in-8° de 163 pp.

45. RIKABI SUCCARI O., International Petroleum Market-Policy Confrontation of the Common Market and the Arab Countries. Louvain, Imprimerie Offset Frankie, 1968. Un vol. in-8° de 224 pp.

46. LEROY R., Essai sur la population active. Théories économiques récentes et analyse régionale de l'emploi féminin. Louvain, Imprimerie Universitaire, 1968. Un vol. in-8° de 224 pp.

47. BARREA J., L'intégration politique externe. Notion globale et analyse sociologique. Paris-Louvain, Editions Béatrice-Nauwelaerts, 1969. Un vol. in-8° de 332 pp.

48. QUERINI G., Agricoltura E sviluppo Economico il caso della Somalia. Roma, Edizioni Ricerche 1969. Un vol. in-8° de 257 pp.

49. VERHOEVEN J., De sociologische verklaring van de sociale controle. Mechelen, Uitgeverij Sinfra, 1969. Un vol. in-8° de 305 pp.

50. DE LANOO I., Stratifikatieproblemen en demokratisering van het universitair onderwijs. Antwerpen, de Nederlandsche Boekhandel, 1969. Un vol. in-8° de 431 pp.

51. QUAETHOVEN P., Het statuut van de ziekenhuisgeneesheer in de Lid-Staten van de Europese Economische Gemeenschap. Kessel-Lo, Uitgeverij van der Poorten, 1969. Un vol. in-8° de 478 pp.

52. DUQUESNE de la VINELLE L., Une théorie des systèmes économiques. Bruxelles, Editions de la Librairie Encyclopédique, 1969. Un vol. in-8° de 345 pp.

53. POURBAIX C., Valeur de l'entreprise, critères et mesures. Paris, Dunod, 1969. Un vol. in-8° de 318 pp.

54. TERMOTE M., Migration et équilibre économique spatial. Louvain, Département Travail, 1969. Un vol. in-8° de 252 pp.

55. VAN OUTRYVE L., Sociologie en vankbond. De vankbond participatie in onderscheiden sociologische modellen. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1969. Un vol. in-8° de 474 pp.

56. VAN GORP G., Le recrutement et la formation des candidats officiers de carrière à l'armée belge. Louvain, 1969. Un vol. in-8° de 554 pp.

57. DESCY J., Critères de rationalisation du développement hospitalier. Louvain, Les Editions André Dewallens, 1969. Un vol. in-8° de 290 pp.

58. JUCKLER F., Modèles de gestion des stocks et coûts marginaux. Louvain, Vander, 1969. Un vol. in-8° de 224 pp.

59. VEZINA P., La détermination du profit comptable et les changements dans les prix. Québec, Les Presses de l'Université Laval, 1969. Un vol. in-8° de 373 pp.

60. COMELIAU C., Conditions de la planification du développement - l'exemple du Congo. Paris- La Haye, Mouton, 1969. Un vol. in-8° de 369 pp.

61. COBBAUT R., La politique de dividende des entreprises belges et américaines 1946-1965. Bruxelles, la Renaissance du Livre, 1969. Un vol. in-8° de 274 pp.

62. BLOES R., Le plan "Fouchet" et le problème de l'Europe politique. Bruges, Collège de l'Europe, 1969. Un vol. in-8° de 538 pp.

63. LOHISSE J., La communication anonyme. Paris, Editions universitaires, 1969. Un vol. in-8° de 198 pp.

64. VAN ROMPUY P., De vraag naar arbeid. Een econometrische studie van conjuncturele tewerkstellingsfunkties. Antwerpen, Standaard Wetenschappelijke Uitgeverij, 1969. Un vol. in-8° de 200 pp.

65. D'OLIESLAGER G., Arbeiders en cultuur. Een sociologische studie over de culturele beeldvorming en de culturele gedragspatronen van arbeiders in de vrije tijd. Antwerpen, Nederlandse Boekhandel, 1969. Un vol. in-8°.

66. DOOGHE G., De struktuur van het gezin en de sociale relaties van de bejaarden. Antwerpen, Nederlandse Boekhandel, 1969. Un vol. in-8°.

67. SIAENS A., Structure des épargnes et conjoncture des capitaux. Louvain, Imprimerie Offset Frankie, 1970. Un vol. in-8° de 276 pp.

68. GERARD H., Catholicisme et fécondité - Recherche exploratoire. Louvain, Vander, 1970. Un vol. in-8° de 230 pp.

69. HUYBRECHTS A., Transports et structures de développement au Congo - Etude du progrès économique de 1900 à 1970. Paris - La Haye, Mouton, 1970. Un vol. in-8° de 418 pp.

70. SAINT-PIERRE J., Théorie économique et financement de la firme. Louvain, Imprimerie Offset Frankie, 1970. Un vol. in-8° de 239 pp.

71. NGUYEN thi THANH-TUYEN, Essai d'une technique projective en sociologie de la religion. Le jeu de construction. Bruxelles, Lumen Vitae, 1970. Un vol. in-8° de 347 pp.

72. LOWENTHAL P., Conjoncture et action des entreprises. Louvain, Centre de Recherches économiques, 1970. Un vol. in-8° de 197 pp.

73. DUPRIEZ P., Contrôle des changes et structures économiques Congo 1960-1967, Paris-La Haye, Mouton, 1970. Un vol. in-8° de 197 pp.

74. de SCHREVEL M., Les forces politiques de la décolonisation congolaise jusqu'à la veille de l'indépendance. Louvain, Imprimerie M. et L. Symons, 1970. Un vol. in-8° de 507 pp.

75. CROTEAU O., La comptabilité, méthode de sélection, de mesure et de communication. Montréal, Les Presses HEC et Dunod Québec Inc., 1971. Un vol. in-8° de 260 pp.

76. SUBEI A.R., L'évaluation économique des projets d'investissement public : le cas du projet de l'Euphrate en Syrie. Louvain, Imprimerie Offset Frankie, 1970. Un vol. in-8° de 208 pp.

77. VANHAEVERBEKE A., Rémunération du travail et commerce extérieur. Essor d'une économie paysanne exportatrice et termes de l'échange des producteurs d'arachides au Sénégal. Louvain, Centre de Recherches des pays en développement, 1970. Un vol. in-8° de 253 pp.

78. COUVREUR J.-P., La décision d'investir et la politique de l'entreprise. Bruxelles, La Renaissance du Livre, 1970. Un vol. in-8° de 306 pp.

79. KAZADI wa DILE J., Politiques et techniques de rémunération dans l'entreprise au Congo. Leurs implications quant au développement économique. Kinshasa - Paris, IRES-Editions Universitaires, 1970. Un vol. in-8° de 482 pp.

80. HUYSE L., Paciviteit-pacificatie en verzuiling in de belgische politiek - een sociologische studie. 1970. Un vol. in-8°.

81. VERCAUTEREN P., Les sous-prolétaires. Essai sur une forme de paupérisme contemporain. Bruxelles, Les Editions Vie Ouvrière, 1970. Un vol. in-8° de 201 pp.

82. DE HAAN H., De toepassing van de dimensie analyse op de econometrie Amsterdam, The North Holland Publishing Company, 1970. Un vol. in-8° de 217 pp.

83. DE DONCKER M., Winstbesteding, groei en aandelenkoersen. Een onderzoek betreffende de Naamloze Venootschappen in Belgïe (1949-1965). 1970. Un vol. in-8° de 418 pp.

84. LACROIX R., Pour une théorie de l'investissement direct étranger dans l'industrie manufacturière. Louvain, Vander, 1970. Un vol. in-8° de 268 pp.

85. ISMAIL T., L'industrie dans l'économie de l'Afrique centrale - Sénégal - Côte d'Ivoire - Cameroun - Congo (Kinshasa). Louvain, Imprimerie Offset Frankie, 1970. Un vol. in-8° de 249 pp.

86. MILLER R.E., Innovation, Organization and Environment - A study of Sixteen American and West European Steel Firms. Sherbrooke, Institut de recherche de perfectionnement en administration, Université de Sherbrooke, 1971. Un vol. in-8° de 211 pp.

87. MORALES J.A., Bayesian Full Information Structural Analysis with an Application to the Study of the Belgian Beef Market. Berlin, Springer-Verlag, 1971. Un vol. in-8° de 154 pp.

87(bis) DESAEYERE W., De vraag naar hoger onderwijs, een dynamisch model. Antwerpen-Utrecht, Standaard Wetenschappelijke Uitgeverij, 1971. Un vol. in-8° de 485 pp.

88. RIKABI-SUCCARI-HART E., Regional System and Political Development in the Arab World. Louvain, Imprimerie Offset Frankie, 1971. Un vol. in-8° de 219 pp.

88(bis) BRONSARD C., Qualité microéconomique et théorie du second best. Louvain, Vander, 1971. Un vol. in-8° de 174 pp.

89. APPELT H., Die Rolle der Geschäftsbanken im Rahmen der Investitionsfinanzierung in der bundesrepublik Deutschland von 195O-1967. Bad Windsheim, Verlag für Wissenschaft und Leben, 1971. Un vol. in-8° de 227 pp.

90. RONG-I WU., The Strategy of Economic Development. A Case Study of Taïwan. Louvain, Vander, 1971. Un vol. in-8° de 217 pp.

91. SALVAS-BRONSARD L., Les techniques quantitatives de la politique économique. Louvain, Vander, 1972. Un vol. in-8° de 211 pp.

92. KALALA M., Politique financière de l'entreprise et conjoncture. Louvain, Imprimerie Offset Frankie, 1972. Un vol. in-8° de 310 pp.

93. JAUMIN C., L'introduction de nouvelles bases techniques en assurance sur la vie. Adaptation du portefeuille en cours. Louvain, Imprimerie Ceuterick, 1972. Un vol. in-8° de 96 pp.

94. SADHU A.N., Surplus Labour and Indian economic development. Vander, 1972. Un vol. in-8° de 186 pp.

94(bis) NILS A., Les radioéléments, facteurs de progrès économiques. Dunod, Paris, 1965. Un vol. in-8° de 208 pp.

95. HERREGAT G., Managerial Profiles and Investment Patterns, an analysis of the International Diffusion of the Basic Oxygen Steel Process. Louvain, Centre de Recherches économiques, 1972. Un vol. in-8° de 191 pp.

95(bis)LEHMANN J.T., La mesure des temps alloués. Librairie Universitaire, 1965. Un vol. in-8° de 335 pp.

96.BARTOLOMEI J.A., Balance des paiements et politique monétaire : L'expérience allemande entre 1958 et 1968. Louvain, Vander, 1972. Un vol. in-8° de 107 pp.

96(bis) SIOUTI M., De l'importance du crédit et des banques dans l'économie. Imprimerie du Gouvernement DAMAS, 1965. Un vol. in-8° de 287 pp.

97. HANQUET H., L'activité professionnelle des femmes : moyen de participation au développement global. Analyse du cas belge. Bruxelles, Les Editions Vie Ouvrière, 1972. Un vol. in-8° de 405 pp.

97(bis) ZAKENIA S., Sur un régime particulier d'assurance sociale. Librairie Universitaire, 1966. Un vol. in-8° de 66 pp.

98. BALLEGEER Y., Approximation de fonctions de poisson généralisées par une fonction de sinistres extrêmes. Louvain, Imprimerie Offset Frankie, 1972. Un vol. in-8° de 110 pp.

99. PUTZ D., Une modélisation de la gestion de la trésorerie. Louvain, Imprimerie Offset Frankie, 1972. Un vol. in-8° de 291 pp.

100. DE SANTIS V., La liberta' e l'uguaglianza nel pensiero di Alexis de Tocqueville. Louvain, Imprimerie Offset Frankie, 1972. Un vol. in-8° de 294 pp.

101. ZAKALNYCKYI W., Attitude des migrants journaliers à l'égard de l'emploi en région. Bruxelles, Les Editions Philippe Charlier, 1973. Un vol. in-8° de 179 pp.

102. MOUCHART M., Analyse Bayesienne du modèle de régression avec résidus non-standards. Louvain, Bibliothèque de l'Université, 1973. Un vol. in-8° de 67 pp.

103. HEES M., Automatisation administrative et qualification. Louvain, Imprimerie Offset Frankie, 1973. Un vol. in-8° de 220 pp.

104. STRABRYLA-THIEL A., La presse quotidienne à la recherche de l'objectivité. Réflexions sur les distorsions de l'information journalistique. Louvain, Imprimerie Offset Frankie, 1973. Un vol. in-8° de 226 pp.

105. DUBOIS L., Un modèle de choix des investissements et des sources de financement. Bruxelles, La Renaissance du Livre, 1973. Un vol. in-8° de 315 pp.

106. KOULO A., Programmation économique et comptabilité opérationnelle. Louvain, Imprimerie Offset Frankie, 1973. Un vol. in-8° de 454 pp.

107. VOYE L., Sociologie du geste religieux. De l'analyse de la pratique dominicale en Belgique à une interprétation théorique. Bruxelles, Les Editions Vie Ouvrière, 1973. Un vol. in-8° de 309 pp.

108. WTTERWULGHE R., L'offre publique d'acquisition au service d'un marché des sociétés. Une analyse droit-économie. Bruxelles, La Renaissance du Livre, 1973. Un vol. in-8° de 281 pp.

109. THIRY J.P., Théories sur le phénomène urbain. Analyse critique et matériaux pour une théorie sociologique. Bruxelles, Imprimerie P. François S.P.R.L., 1973. Un vol. in-8° de 253 pp.

110. HEYVAERT H., Stratégie et innovation dans l'entreprise. Louvain, Imprimerie CIACO, 1973. Un vol. in-8° de 266 pp.

111. BELLEHUMEUR A., La quasi-monnaie. Québec, Imprimerie Fernand ENR, 1973. Un vol. in-8° de 356 pp.

112. HOUYOUX J., Budgets ménagers, nutrition et mode de vie à Kinshasa (République du Zaïre). Kinshasa, Presses Universitaires du Zaïre, 1973. Un vol. in-8° de 303 pp.

113. MARTOU F., Stratégie et équilibre des ressources humaines. Louvain, Imprimerie Offset Frankie, 1973. Un vol. in-8° de 181 pp.

114. WILMES P., Un modèle de transport multi-ressources. Application de la dynamique des systèmes. Louvain, Imprimerie Offset Frankie, 1973. Un vol. in-8° de 358 pp.

115. ALETUM M.T., Political Conflicts within the Traditional and the Modern Institutions : A Case Study of the Bafut-Cameroun. Louvain, Vander, 1973. Un vol. in-8° de 175 pp.

116. KUMPS A.-M., Le conglomérat, nouvelle forme de concentration. Le cas anglais. Bruxelles, La Renaissance du Livre, 1974. Un vol. in-8° de 228 pp.

117. RICHTER F., Les cadres formés à l'étranger dans la société sud-vietnamienne. Louvain, 1974. Un vol. in-8° de 256 pp.

118. LEBLANC G., Trois applications économiques de la programmation convexe duale. Louvain, Vander, 1974. Un vol. in-8° de 64 pp.

119. HERNANDEZ LOPEZ M., Les déterminants des taux d'intérêt - discussion théorique et estimation empirique - application au cas belge. Louvain, CIACO, 1975. Un vol. in-8° de 276 pp.

120. COLARD M., Dynamique du marché et comportement d'achat. Louvain Center for Socio-Economic Studies in Advertising and Marketing - Institut d'administration et de gestion, 1975. Un vol. in-8° de 169 pp.

121. PARK J.Y.C., Progrès, prix et coûts - Etude de diffusion des progrès de convergence des prix, Japon - Etats-Unis, 1868-1970. Louvain, Imprimerie Offset Frankie, 1975. Un vol. in-8° de 277 pp.

122. DECALUWE B., Le régime du double marché des changes - Théorie et pratique - L'expérience de l'union économique belgo-luxembourgeoise. Mecaprint, rue F. Severin, 24, Bruxelles, 1975. Un vol. in-8° de 261 pp.

123. ROUZIER P., The Evaluation of Optimal Monetary and Fiscal Policies with a Macroeconomic Model for Belgium. Faculté des Sciences économiques, sociales et politiques, Louvain, 1975. Un vol. in-8° de 156 pp.

124. SHABANI NYEMBO, L'industrie du cuivre dans le monde et le progrès économique deu copperbelt africain. La Renaissance du Livre, Bruxelles, 1975 . Un vol. in-8° de 270 pp.

125. BUEHLER A., L'innovation dans l'institution religieuse - Une analyse sociologique de trois groupes de prêtres contestataires : France, Allemagne, Hollande. U.C.L., Faculté des Sciences économiques, sociales et politiques. Un vol. in-8° de 405 pp.

126. WARNOTTE G., Processus d'évaluation et logique organisationnelle. Faculté des Sciences économiques, sociales et politiques, Louvain, 1975. Un vol. in-8° de 230 pp.

127. CLOSON J.-P., Santé publique et prévoyance sociale. Une approche économique. Faculté des Sciences économiques, sociales et politiques, Louvain, 1975. Un vol. in-8° de 105 pp.

128. DESCHAMPS J.-P., Second-best Pricing with Variable Product Quality: a Quantitative Decision Rule and Existence Proof. Louvain, 1976. Un vol. in-8° de 73 pp.

129. WOICHE M., La modélisation en gestion intégrée à court terme. Application de procédures de prise de décision décentralisée. Louvain, 1976. Un vol. in-8° de 242 pp.

130. TWAGIRAMUTARA P., Rapport à l'espace, rapport social. Contribution à l'étude de leurs liaisons à partir du cas du Rwanda. Louvain, 1976. Un vol. in-8° de 246 pp.

131. JULIEN P.A., Planification et domination au Québec. 1977. Un vol. in-8°.

132. ANDRE J., Coût du capital et théorie du financement de l'entreprise. 1977. Un vol. in-8° de 382 pp.

133. BODSON P., Une dynamique du marché du travail. 1977. Un vol. in-8° de 367 pp.

134. RIGAUX-BRICMONT B., La prise de décisions économiques dans la famille. Louvain-la-Neuve, Center for Socio-Economic Studies in advertising and marketing, Institut d'administration et de gestion, 1977. Un vol. in-8° de 288 pp.

135. AL YOUSEF M., Analyse de modèles de risques compétitifs. 1977. Un vol. in-8°.

136. BLOMME R., Etude des propriétés dynamiques d'un modèle stochastique : application au cas modèle anelise. 1978. Un vol. in-8° de 295 pp.

136(bis) KABAMBA H., Le crédit commercial des banques d'émission et l'attitude de l'autorité monétaire à son égard. Institut des Sciences économiques, 1978. Un vol. in-8° de 141 pp.

137. de CARVALHO F.A., Planning public consumption under restricted information, a process-oriented contribution. Louvain-la-Neuve, CIACO, 1979. Un vol. in-8° de 149 pp.

138. LACONTE P., Mutations urbaines et marchés immobiliers : le développement des immeubles de bureaux à Bruxelles. Louvain, OYEZ, 1978. Un vol. in-8° de 302 pp.

139. HERMAN J., L'articulation des sciences et l'organisation de la recherche. Transdisciplinarité, paradigmes méthodologiques et socio-genèse des pratiques scientifiques. Louvain, Imprimerie Offset Frankie, 1978. Un vol. in-8° de 349 pp.

140. SCHMIEGELOW M., System charge or system maintenance ? Problems of functionalism in european monetary integration. Sri-Lanka, Imprimerie Gunasena et Cie, 1979. Un vol. in-8° de 162 pp.

141. GERNAY C., Planification stratégique et discontinuité. Application au cas des compagnies aériennes. CRECIS Centre de recherches : Changement, Innovation, Stratégie. Un vol. in-8° de 404 pp.

142. VAN NAMEN P., La gestion d'un portefeuille d'assurances R.C. Automobile. Bruxelles, Office des Assureurs de Belgique, 1979. Un vol. in-8° de 120 pp.

143. QUINTART A., La décision de désinvestir dans l'entreprise. CREFIM, Centre de recherche en gestion financière, Institut d'administration et de gestion, 1979. Un vol. in-8° de 446 pp.

144. KIONI KIABANTU T., Procédures dynamiques d'allocation des ressources dans l'espace. Une application à la planification des infrastructures de transport. 1980. Un vol. in-8° de 166 pp.

145. ILMANE M.Ch., Internationalisation du capital et industrialisation du Tiers-Monde. Un essai de reformulation de la problématique de sous-développement à travers l'exemple algérien. Louvain-la-Neuve, Institut d'administration et de gestion, 1981. Un vol. in-8° de 337 pp.

146. SILVA LEITON R., Le follow-up de la décision d'investir. CREFIM, Centre de recherches en gestion financière, Institut d'administration et de gestion, Louvain-la-Neuve, 1981. Un vol. in-8° de 274 pp.

147. MALEVEZ P., Modélisation de la gestion bilantaire à court terme d'une banque. CRECIS, Centre de recherches : Changement, Innovation, Stratégie, Institut d'administration et de gestion, Louvain-la-Neuve, 1981. Un vol. in-8° de 375 pp.

148. CHUNG J.W., La théorie de la baisse tendancielle du taux de profit : analyse théorique et application empirique du développement du capitalisme américain 1929-1978. Economie et Société (ECOS), Institut des sciences économiques, Louvain-la-Neuve, 1981. Un vol. in-8° de 152 pp.

149. MELESE d'HOSPITAL F., The optimal management of nonrenewable energy resources : problems of entry and taxation. Louvain-la-Neuve, 1981. Un vol. in-8° de 130 pp.

150. VEGA-CENTENO M., Naturaleza y caracteristicas del cambio técnico en un Pais sub-desarrollado : el caso de la Industria Manufacturera en el Peru. Louvain-la-Neuve, CIACO, 1982. Un vol. in-8° de 293 pp.

151. GAMELA NGINU DIAMUAN GANA Performances de l'entreprise dans les pays en voie de développement : le rôle de la fonction approvisionnement dans la stratégie de l'entreprise (cas des entreprises manufacturières du Zaïre). Louvain-la-Neuve, Cabay Libraire-Editeur, 1982. Un vol. in-8° de 313 pp.

152. MUNKENI LAKUP TIER, Le recours à des formes fonctionnelles flexibles dans l'estimation des systèmes de demande et dans le test d'hypothèse néo-classique. 1983. Un vol. in-8° de 174 pp.

153. DE LECEA Antonio, On fiscal Federalism, Information and Regional Competition. Louvain-la-Neuve, Cabay, 1983. Un vol. in-8° de 234 pp.

154. CHOU Tein Chen, Industrial Organization in the Process of Economic Development. The case of Taïwan 1950-1980. Louvain-la-Neuve, CIACO, 1985. Un vol. in-8° de 158 pp.

155. LAKSACI Mohammed, Financement monétaire de l'investissement productif. Application de l'analyse au cas algérien. Louvain-la-Neuve, CIACO, 1985. Un vol. in-8° de 235 pp.

156. LAMBERT J.-P., Disequilibrium macro models based on business survey data. Theory and estimation for the Belgian manufacturing sector. Louvain-la-Neuve, 1984. Un vol. in-8° de 187 pp.

157. WIBAUT Serge, Tax reform in desequilibrium economies : an application to Belgium (1982). Louvain-la-Neuve, 1986. Un vol. in-8° de 132 pp.

158. TSHIMANGA Bakadia-Babu, Transferts de technologie dans les pays en développement : le secteur agro-alimentaire au Zaïre. Louvain-la-Neuve, 1986. Un vol. in-8° de 343 pp.

159. FONTENELE BESSA C.A., Le comportement d'achat des ménages en milieu urbain. Louvain-la-Neuve, CIACO, 1986. Un vol. in-8° de 364 pp.

160. WYLLEMAN Edouard, On the use of business survey information in macroeconometric models. Louvain-la-Neuve, CIACO, 1986. Un vol. in-8° de 128 pp.

161. VERA Hector, Le rôle idéologique des mass-media en Amérique latine. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8° de 100 pp.

162. ORSI Renzo, Une analyse économétrique du marché du travail dans l'industrie manufacturière italienne. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8° de 113 pp.

163. GHOUFI Abdelhamid, Choix de politique économique et efficacité dans l'allocation des ressources. Une analyse globale et semi-globale de l'investissement en Algérie. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8° de 230 pp.

164. GARELLA Paolo, Imperfect information in Markets with few competitors. Louvain-la-Neuve, CIACO, 1986. Un vol. in-8° de 56 pp.

165. MPUNGU MULENDA Saidi, Un regard en marge. Le public populaire du cinéma au Zaïre. Louvain-la-Neuve, 1987. Un vol. in-8° de 166 pp.

166. SY Malick, Simultaneous Expiration Dates Effects in Index Futures and Options : Arbitrage and Hedging. Louvain-la-Neuve, 1987. Un vol. in-8° de 303 pp.

167. AJA HAMID, Le rôle du marketing dans la politique de l'entreprise hôtelière. Application au cas irakien. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8° de 424 pp.

168. STEEL Mark, Bayesian Analysis of multivariate Exogeneity : a Monte-Carlo Approach. Louvain-la-Neuve, 1987. Un vol. in-8° de 125 pp.

169. CREMER Helmuth, Essays in Public Economics. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8°.

170. INGHAM John, Evaluation dynamique des systèmes d'information de gestion. Louvain-la-Neuve, ACADEMIA Edition et Diffusion, 1987. Un vol. in-8° de 380 pp.

171. MODESTO Maria Leonore Martins Ribiero PRICE, PRODUCTION AND INVENTORIES ADJUSTMENTS: MICROFOUNDATIONS AND AGGREGATE BEHAVIOUR. An Empirical Study for the Belgian Manufacturing Sector. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8° de 198 pp.

172. LUZ Rogerio, ESPACE POTENTIEL ET EXPERIENCE FILMIQUE. Une application des propositions de D.W. Winnicott concernant la culture. Louvain-la-Neuve, CIACO, 1987. Un vol. in-8° de 405 pp.

173. MARTINEZ-GIRALT Xavier, Essays in the economics of product variety. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8°.

174. LOPEZ GARCIA Salvador, A normative analysis on wage subsidies and unemployment benefits.

175. KEYMOLEN Guy, Les problèmes de tournées : un algorithme pour le cas monovéhicule avec restrictions temporelles et de précédence. Louvain-la-Neuve, ACADEMIA Edition et Diffusion, 1988. Un vol. in-8° de 287 pp.

176. PACHECO Fernando, Economic Policy in Open Economy.

177. ANTOINE Frédéric, De la télévision de l'imaginaire à l'imaginaire télévisuel. La télévision à travers ses programmes. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 414 pp.

178. DA COSTA GARCIA Manuel, Education et niveau de vie - Le cas du Portugal. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 80 pp.

179. SARQUES Jane Jorge, L'écran magique - Les émissions pour enfants et leur impact culturel (Brésil). Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 271 pp.

180. TSHISHI BAVUALA MATANDA, Littérature et cinéma. Contribution à l'étude de l'adaptation dans la culture négro-africaine. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 178 pp.

181. GOLLIER Christian, Intergenerational risk-sharing and unemployment. Louvain-la-Neuve, 1988. Un vol. in-8° de 192 pp.

182. YU Hua, Commercial Bank's Sovereign Loan Lending Behavior and Loan Portfolio Analysis. Louvain-la-Neuve, ACADEMIA Edition et Diffusion, 1989. Un vol. in-8° de 173 pp.

183. LUSENDI MATUKAMA, Les déterminants de la répartition des bénéfices des entreprises au Zaïre. Une approche économétrique. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 182 pp.

184. CHAU Fernando, Essays on credit contracts. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 122 pp.

185. CARRASCO GAETE Luis Erico, Le pouvoir et la presse hebdomadaire. Essai d'analyse pragmatique. Louvain-la-Neuve, CIACO, 1988. Un vol. in-8° de 289 pp.

186. DEPRINS Dominique, Estimation de frontières de production et mesures de l'efficacité technique. Louvain-la-Neuve, CIACO, 1989. Un vol. in-8° de 84 pp.

187. LEGROS Edmond, L'ANALYSE DES CORRESPONDANCES. Démarches et réflexions critiques. Louvain-la-Neuve, CIACO, 1989. Un vol. in-8° de 147 pp.

188. KASONGO IBANDA NGOZULU W.T., Le "cinéma pour Africains", acte d'influence à visée persuasive. Louvain-la-Neuve, CIACO, 1989. Un vol. in-8° de 193 pp.

189. ZOU Liang, Essays in Principal-Agent Theory. Louvain-la-Neuve, CIACO, 1989. Un vol. in-8° de 154 pp.

190. BEER Francisca, The Existence of the Dividend Signaling Equilibria : an Analysis of the Belgian Case. Louvain-la-Neuve, CIACO, 1990. Un vol. in-8° de 216 pp.

191. LEFEBVRE Olivier, Sécuritisation et Intermédiation Financière de Dépôt. Louvain-la-Neuve, CIACO, 1990. Un vol. in-8°.

192. LOPEZ Salvador, Wage Subsidies in Disequilibrium Economies. Louvain-la-Neuve, CIACO, 1990. Un vol. in-8° de 139 pp.

193. MPOYI MUKENDI, La communication politique en Afrique. Une approche pragmatique. Louvain-la-Neuve, CIACO, 1990. Un vol. in-8° de 236 pp.

194. de BRIER Claire, Les pratiques de formation dans les entreprises belges. Louvain-la-Neuve, 1990. Un vol. in-8° de 367 pp.

195. ALBAEK Svend, Essays in the Theory of Duopoly with Incomplete Information. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 150 pp.

196. MULKAY Benoît, Investment and Business Survey. A Study on a Panel of French Industrial Firm. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 390 pp.

197. LICANDRO Omar, Investment and The Stock Market in Quantity Rationing Models. Louvain-la-Neuve, CIACO, 1990. Un vol. in-8° de 138 pp.

198. DEFOURNY Vincent, Organisation et pilotage stratégique de la communication. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 256 pp.

199. KAZADI Tshamala, Les déterminants de la crise agraire en Afrique subsaharienne, la spécificité du Zaïre. Rôle du modèle de développement et des agents socio-économiques. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 241 pp.

200. DE RONGE Yves, Analyse du processus de filialisation d'entreprises multidivisionnelles : approche théorique et étude empirique sur le marché belge. Louvain-la-Neuve, 1991. Un vol. in-8° de 357 pp.

201. GILLET Roland, L'efficience informationnelle du marché boursier : aspects théoriques et empiriques. Louvain-la-Neuve, 1991. Un vol. in-8° de 148 pp.

202. ARAVENA NAVARRETE José, Debt Reduction Schemes. Theoretical Issues and Empirical Results for Chile. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 253 pp.

203. DUARTE NEVES P., Commodity Demands and Female Labour Supply over the Life-Cycle. Louvain-la-Neuve, 1991. Un vol. in-8° de 191 pp.

204. MARION Philippe, Traces en cases.
Travail graphique, figuration narrative et participation du lecteur. Essai sur la bande dessinée.

205. OVERGAARD Per Baltzer, PRODUCT QUALITY UNCERTAINTY. Strategic Information Transmission in Product Markets with Adverse Selection and Adverse Incentives. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 273 pp.

206. LEBRUN Bernard, Asymmetry in Auctions. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 274 pp.

207.YEN Ping-Yuan, Essays on One-Point Solution Concepts for The Provision of a Public Good. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 89 pp.

208. MARQUET Jacques, Nomisation et réalité dynamiques. Contribution à la sociologie compréhensive.Louvain-la-Neuve, Academia Edition & Diffusion, 1991. Un vol. in-8° de 284 pp.

209. MOTTA Massimo, Strategic Investments, R & D and International Economics. Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 119 pp.

210. GONZALEZ Jorge, La politique fiscale et budgétaire en Colombie (1964-1988). Louvain-la-Neuve, CIACO, 1991. Un vol. in-8° de 202 pp.

211. de BODT Eric, Le traitement de l'information marketing et financière dans la décision d'investissement. Une intégration par les systèmes experts. Louvain-la-Neuve, 1992. Un vol. in-8° de 402 pp.

212. DESCLEE de MAREDSOUS Xavier,
Diversification et performance économique : le facteur "Organisation".
Analyse comparée de deux Groupes belges diversifiés. Louvain-la-Neuve, 1992. Un vol. in-8° de 342 pp.

213. BARHAM Vicky, Issues in Club Theory and Social Equity. Louvain-la-Neuve, CIACO, 1992. Un vol. in-8° de 138 pp.

214. DOR Eric, Modèles de consommation avec contraintes quantitatives et financières. Louvain-la-Neuve, CIACO, 1992. Un vol. in-8° de 262 pp.

215. WOLA-MBALE IMBOPO,
Analyse des projets industriels dans les entreprises zaïroises (1970-1987) - La décision d'investissement et de financement.

216. RODRIGUEZ POO Juan M.,
Constrained Nonparametric Regression. Louvain-la-Neuve, CIACO, 1992. Un vol. in-8° de 194 pp.

217. YEPEZ DEL CASTILLO Isabel,
Les syndicats à l'heure de la précarisation de l'emploi. Une approche comparative Europe-Amérique latine au départ d'une étude de cas au Pérou. Louvain-la-Neuve, Academia, 1993. Un vol. in-8° de 327 pp.

218. BUGHIN Jacques, Union-Firm Bargaining and Imperfect Product Competition in Belgian Manufacturing. Louvain-la-Neuve, CIACO, 1992. Un vol. in-8° de 231 pp.

219. de SOUSA RAMOS Francisco,
Modèles statiques et dynamiques d'optimisation d'une économie avec pollution. Louvain-la-Neuve, CIACO, 1992. Un vol. in-8° de 154 pp.

220. LACROSSE Jean-Marie,
L'efficacité symbolique dans les psychothérapies.

221. BARROS Fatima, The Design of Incentive Schemes in Oligopoly Theory. Louvain-la-Neuve, CIACO, 1993. Un vol. in-8° de 128 pp.

222. DEREZE Gérard, Une ethnosociologie des objets domestico-médiatiques. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 296 pp.

223. GERMAIN Marc, Essais de dynamique économique : changements qualitatifs et apprentissage dans un contexte keynésien. Louvain-la-Neuve, CIACO, 1993. Un vol. in-8° de 146 pp.

224. GREGOIRE Philippe, Gestion d'un portefeuille obligataire et habitat préféré.

225. CORDELLA Tito, Essays in International Trade and Strategic Behaviour. Louvain-la-Neuve, CIACO, 1993. Un vol. in-8° de 74 pp.

226. GUTIERREZ Luiz Homero Silva,
Les transformations de la fonction personnel. La contribution de l'analyse contextualiste. (3 études de cas réalisées au Brésil).

227. GRILO Isabel, Essays on Price Competition and Product Differentiation. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 104 pp.

228. CODOGNATO Giulio, Cournot-Walras Equilibrium in Exchange Economies. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 62 pp.

229. YAPU Mario, L'organisation des savoirs scolaires dans l'enseignement technique et professionnel en Belgique francophone. Louvain-la-Neuve, CIACO, 1994.

230. LEONARD Evelyne, Management des ressources humaines : pour une lecture critique. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 293 pp.

231. NYSSENS Marthe, Quatre essais sur l'économie populaire urbaine : le cas de Santiago du Chili. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 193 pp.

232. PIQUARD Brigitte, Le leadership à l'âge de l'angoisse. Analyse des perceptions du leadership de Zia-Ul-Haq et Benazir Bhutto, à travers des interviews d'intellectuels organiques. Analyse du cas du Pakistan. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 376 pp.

233. BRIANZA Tiziano, Essays on Futures Markets and Imperfect Competition. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 97 pp.

234. SHIKAYI Luboya, Initiatives de développement local et pouvoir paysan. Etude d'une dynamique locale de développement : Le rôle des Associations villageoises, des Organisations non-gouvernementales et des Micro-projets dans la Vallée de la Ruzizi (1978-1989). Sud-Kivu / Zaïre. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 281 pp.

235. DE CARLO Manuela, La gestion des joint-ventures entre concurrents dans les secteurs en phase de maturité. Stabilité et performance des stratégies de coopération. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 276 pp.

236. LOKOTA EKOT'E Panga,
La crise de l'industrialisation, la problématique des activités informelles et les perspectives du développement endogène en Afrique subsaharienne. Cas du Zaïre. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 362 pp.

237. MONFORT Philippe, Export Pricing Behavior, Foreign Trade Performance and Capacity Constraint. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 188 pp.

238. OKITUNDU AVOKI René,
Les petites organisations communautaires comme espace de développement local et base de la maîtrise du processus du développement national. Perspectives et limites. Cas de la Communauté de Base au Zaïre. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 269 pp.

239. RODRIGUEZ HERRERA Adolfo,
Le travail et la formation des prix. Louvain-la-Neuve, CIACO, 1994. Un vol. in-8° de 208 pp.

240. FRANCO Giovanni, Vers un modèle systémique de la structure de l'économie capitaliste. Eléments de traduction systémique du "Capital de Marx".

241. WAUTELET Jean-Marie,
Structures industrielles et reproduction élargie du capital en Belgique - 1850-1914. Louvain-la-Neuve, Academia, 1995. Un vol. in-8° de 417 pp.

242. BIANCHI Marco, Time Series Modelling in the Presence of Structural Change. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 126 pp.

243. RIVERA CAMINO Jaime,
L'orientation marché : une stratégie concurrentielle performante. Application aux secteurs des banques et assurances au Pérou et des assurances particulières en Belgique. Louvain-la-Neuve, CIACO, 1995.

244. DE WINNE Rudy, Structure des taux d'intérêt et processus de diffusion. Une comparaison empirique par inférence indirecte. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 170 pp.

245. MBWINGA Bila, Déterminants et stratégies de compétitivité industrielle en Afrique subsaharienne. Louvain-la-Neuve, Academia, 1995.

246. CROTT Ralph, Evaluation de l'impact des aides publiques directes à la recherche industrielle : une étude empirique sur entreprises wallonnes et belges. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 314 pp.

247. LAPOINTE Marc-André,
Risque, réputation et signalisation par le dividende. Louvain-la-Neuve, CIACO, 1995.

248. SCHNEIDER ARNSPERGER Christian,
Modeling Economic Justice : Five Essays. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 128 pp.

249. GOERGEN Pascal, Le rapatriement des Forces Belges d'Allemagne. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 215 pp.

250. SEMPERE MONERRIS José Jorge, Licensing, Externalities and Competition. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 112 pp.

251. WAUTHY Xavier, Agent's Heterogeneity and Market Outcomes. Some Applications of Addres-Models in Differentiated Markets. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 116 pp.

252. VANNINI Stefano, Essays on the Strategic Behaviour of Multinational Enterprises. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 139 pp.

253. VENTURA Luigi, On the Economic Implications of Incomplete Markets. Louvain-la-Neuve, CIACO, 1995. Un vol. in-8° de 116 pp.

254. SOUGNE Danielle, La consolidation des comptes en Belgique : études empiriques et réflexions théoriques. Louvain-la-Neuve, CIACO, 1995.

255. FAGNART Jean-François, Imperfect Competition, Quantity Constraints and Macroeconomic Dynamics. Louvain-la-Neuve, CIACO, 1995.

256. MINELLI Enrico, Rational Expectations in Games. Louvain-la-Neuve, CIACO, 1995.

257. MONER COLONQUES Rafael Essays in International Oligopoly. Louvain-la-Neuve, CIACO, 1995.

258. PINILLA José Les travailleurs sociaux face à l'exclusion. Pistes novatrices au plan de la formation des professionnels et des politiques sociales. Louvain-la-Neuve, CIACO, 1995.